MULTIPLE REPRESENTATIONS OF CONCEPT FORMATION

By

CARL W. TURNER III

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1993

ACKNOWLEDGMENTS

I wish to express my gratitude to the members of my dissertation committee: Howard Beck, Jeff Farrar, Richard Griggs, C. Michael Levy, and particularly to my advisor Ira Fischler.

I am also deeply grateful to Don Dulany and Russ Poldrack at the University of Illinois for sharing with me computer programs used to validate the statistical analysis of Experiment 1a, and for their patient explanations about the fine points of this analysis.

Thanks go to Heather Howes and Hollie Altman for data collection in Experiments 2 and 3.

My deepest appreciation goes to my wife Krista Thoren, for her emotional and material support during my time as a graduate student.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

MULTIPLE REPRESENTATIONS OF CONCEPT FORMATION

By

Carl W. Turner III

August 1993

Chairperson: Dr. Ira Fischler
Major Department: Psychology

This study addressed the current debate over the relative contributions of recognition memory for category members and abstract knowledge of category structure in classification behavior. Three experiments were conducted using the familiar artificial grammar paradigm in which subjects study rule-ordered letter strings, then make classification judgments of novel letter strings. The experiments, and supporting computer simulations, supported the existence of abstract knowledge in human concept learning. In Experiment 1, subjects classified test items and reported the item fragments that led to their judgments; the experiment replicated an earlier study in which the computed rule validities of reported item fragments closely followed the subjects' percentage of correct judgments. A computer program that simulated simple recognition and report of item fragments in the same task showed that the

correspondence between the two measures could be accounted for by separate computations for feature recognition and item classification. In Experiment 2, modification of the report procedure failed to alter the close correspondence between the computed rules validities and percentage correct judgments, demonstrating the difficulty of dissociating recognition and classification measures in this task. In Experiment 3, subjects participated in one of four groups based on letter set (same or different) and display type (strings or pairs). The experiment included two "no-transfer" trial blocks and a third, "transfer" trial block, in which the test items were constructed of an unfamiliar, unstudied letter set. Subjects in the different-letter-set condition studied novel letter sets in each trial block; those in the same-letter-set condition saw a novel letter set only in the transfer test phase. In the pairs condition, letters were paired for display during study, rather than in a typical whole-item strings display. The different-strings group outperformed the other three groups in both the no-transfer and transfer tests, demonstrating the importance of variable stimuli and whole items during study for developing an abstract representation of the test domain. The research argues for the importance of knowledge of structure both across and within items in the formation of concepts.

CHAPTER 1

INTRODUCTION

The ability of humans to acquire concepts forms the basis of intelligent action. The world's objects and activities could be assigned to categories in an infinite number of ways, presenting humans with a bewildering amount of information. Only a subset is learned; the knowledge of those categories we find useful and necessary is mentally represented as concepts. Concepts allow us to go beyond the information given, to make generalizations, and to apply old knowledge to new situations. "Without concepts, mental life would be chaotic" (Smith & Medin, 1981, p.1).

As in other healthy domains of psychological inquiry, there are many issues in concept formation and categorization still unresolved. Discussions of concept formation are often framed in terms of competing theories; frequently, the question asked is "Given some behavioral data, which of two (or more) theories best accounts of the data?" One such pair of competing theories is the *prototype* and *exemplar* theories. These theories are better described as classes of theories, or perspectives. Another controversy in the literature of concept formation pits theories of *implicit learning and tacit knowledge* against conventional theories of concept formation, referred to here

as *explicit* theories. The present research considers the problem of explaining concept formation from each of these competing views, and attempts to reconcile different approaches rather than confirm or disconfirm any single theory.

Prototype Versus Exemplar Theories

A popular and well-researched issue in the literature of concept formation deals with the distinction between prototype theories and exemplar-based theories. According to prototype theories, concepts are formed through observation and induction of category members, and exist as abstract knowledge, distinct from explicit memory for individual members (Hayes-Roth & Hayes-Roth, 1977; Posner & Keele, 1970; Rosch & Mervis, 1975). Exemplar theories hold that concepts depend only on explicitly remembered category members (Medin & Schaffer, 1978; Nosofsky, 1986; Smith & Medin, 1981). Any fair account of the issue acknowledges the difficulty of distinguishing between prototype and exemplar-based theories, and some suggest that both processes may be at work (Allen & Brooks, 1991; Malt, 1989; McAndrews & Moscovitch, 1985).

Implicit Versus Explicit Learning

Implicit learning has been described as an automatic, effortless process that results in knowledge that is abstract, powerful, and unavailable to conscious awareness (Reber, 1989). According to this hypothesis, implicit knowledge is characterized by the following features: (a)

automatic, or passive acquisition of the structure of complex stimuli; (b) independence from explicit knowledge, available for use in making judgments but not available to conscious awareness; and (c) abstract knowledge, independent from memory for specific episodes or stimuli. This theory in support of a fully cognitive unconscious is contrasted with theories that explain all categorization behavior, judgment, and decision making as purely conscious, fully rational processes (Dulany, 1991).

The issue of the form of knowledge created during concept formation has emerged again in the debate on implicit knowledge (Dulany, Carlson, & Dewey, 1984; Reber, Allen, & Regan, 1985). Broadly stated, we see in this debate many of the concerns raised in the discussions over prototype versus exemplar theories. What is the form of knowledge? Is it possible to dissociate procedural knowledge and verbal awareness in the performance of some task? What constitutes explicit knowledge of contingencies in some stimulus domain? The present research attempted to reconcile some of the findings in implicit learning and tacit knowledge with some of the larger issues in concept formation.

Studies of concept formation frequently rely on artificial stimuli. The instrument used in all of the experiments presented here is the *artificial grammar task*. In a typical artificial grammar experiment, subjects study lines of letters that are ordered by a set of predetermined

rules. In a subsequent test phase, subjects must classify novel lines of letters as well-formed (those believed to follow the rules) or ill-formed (those believed to violate the rules). The process by which humans categorize unfamiliar objects--whether artificial stimuli or natural objects--forms the basis of continued research in learning, memory, and concept formation.

If viewed as a special case in the larger issues over concept formation, the artificial grammar tasks can provide an important contribution to theories of knowledge representation. Likewise, some of the viewpoints generated in the concept formation and categorization literature may shed light on the controversy over implicit learning. This research attempts to place the artificial grammar studies within the context of some unresolved arguments over prototype/exemplar theories of concept formation, and argues for the existence of separate facilities for computing classification decisions and recognition memory.

The results of four experiments are presented. Three experiments involved human subjects making classification judgments in artificial grammar tasks; one involved a computer simulation of an information processing model of the task. Experiment 1a replicated an earlier artificial grammar study which purported to show that subjects' verbal reports could fully account for their classification judgments of letter strings (Dulany et al., 1984). Experiment 1a also provided a set of data upon which a

computational model was developed of recognition and reporting in the task. Experiment 1b reports the details of the model and describes the outcome of a computer simulation.

Experiment 2 modified the design of Experiment 1a in an effort to uncover dissociations between verbal report and classification performance. The experiment also compared the differences between two groups of observation learners and one group which learned the task as an explicit test for memory of logical rules. The lack of success in finding a dissociation between classification performance and two verbal report measures is discussed both in terms of experimental methodology and directions for future research.

Experiment 3 included a novel transfer task in which subjects made classification judgments of letter strings created from sets of letters that did not appear during study. The experiment demonstrated that classification judgments could be made in the absence of explicit recognition memory of studied items; the test items in this case shared no surface features with the studied items. The results were offered in support of the hypothesis that subjects were abstracting the underlying structure of the stimuli both within and across studied items and using knowledge of that structure to correctly classify the novel test items. The results also showed, contrary to some previous research, the advantage of variable, whole-item stimuli during study on the ability to classify novel test

items. The discussion of this experiment centers on the importance of abstract knowledge of structure in concept formation and is tied to the issue of separate processes for classification and recognition of exemplars.

CHAPTER 2

EXPERIMENT 1A

The arguments over the existence of implicit knowledge in artificial grammar tasks are well known. Typically, subjects studied strings of letters ordered by an artificial grammar (see Figure 1).[1] In a subsequent test phase, they were asked to discriminate between well-formed (grammatical) letter strings and ill-formed (nongrammatical) letter strings based on their experience with the items seen during the initial study phase (Reber, 1967, 1976; Reber & Allen, 1978; Reber, Kassin, Lewis, & Cantor, 1980). Implicit learning, it is claimed, occurs automatically and results in abstract knowledge (Reber, 1989). In artificial grammar tasks, this abstract knowledge takes the form of unconscious grammatical "rules" that can be accessed to classify novel stimuli. The most contentious of the claims for implicit knowledge is the assertion that subjects are able to make better-than-chance classifications of these complex stimuli without being able to verbally justify the reasons for their judgments.

Dulany, Carlson, and Dewey (1984) addressed this issue of classification and verbal awareness in the artificial grammar tasks. In a study based on Reber (1976) and Reber and Allen (1978), subjects justified their classification

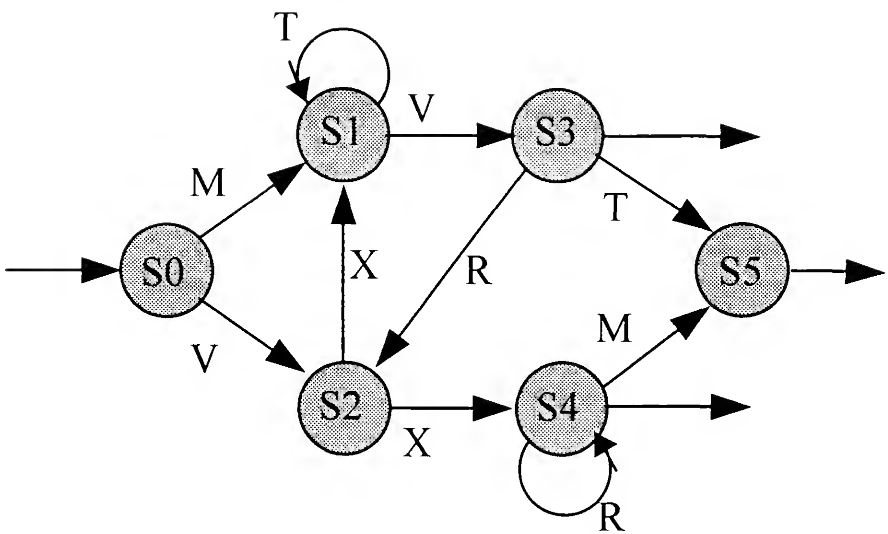

Figure 1. A representation of the finite-state grammar used to generate the grammatical letter strings.

judgments by marking on their scoresheets a fragment of the test item that made the item correct or incorrect. Each item fragment, they argued, constituted a *conscious* "rule" used by the subject to make a judgment. Each of these reported "rules" was associated with a computable "cue validity" based on their occurrence in the set of test items.[2] A regression analysis of the mean rule validities and proportion correct judgments for 50 subjects yielded an intercept of .01 and a slope of .99. Thus, mean rule validity "predicted" the percent correct judgments for a large sample of subjects. In their view, classification judgments were not the result of nonconscious knowledge of grammatical rules, as Reber had insisted; they resulted from imperfectly learned rules, fully available to conscious awareness (Dulany et al., 1984).

These reported "rules" are somewhat different in character than the logical rules that order the letters, as represented by Figure 1: they represent combinations of letters that occur together within letter strings. Previous research had shown that the knowledge acquired from the observation of rule-ordered strings is qualitatively different from that gained by explicit knowledge of the logical rules used to order the letter strings (Turner & Fischler, 1993). Experiment 2 in the present study introduced a training manipulation which contrasted learning by observation with explicit memory for logical rules. Therefore, in order to prevent misunderstanding, reported

item fragments will henceforth be referred to as *featural rules*, in contrast with the knowledge of *logical rules*.

Might the same data have come from subjects correctly classifying strings, then randomly marking combinations of letters within the test items (Reber et al., 1985)? It was hypothesized that subjects simply making random reports after classification judgments would produce the same outcome described in Dulany et al. (1984). To test this hypothesis Reber et al. (1985) generated judgments and featural rule reports for 15 simulated "subjects" and found that the results followed closely those of Dulany et al. Regression analysis yielded a line with a unit slope (slope = .96, intercept = .06, r = .82). Reber et al. suggested that subjects were merely guessing when asked to justify their judgments, and concluded that the close association of computed rule validity and percent correct was an artifact, produced by the constraints of the experimental design.[3]

Dulany, Carlson, and Dewey (1985), in reply, rejected this methodological criticism, insisting that the act of identifying and providing featural rules relied on recall and verbal knowledge. They offered their own computer simulation of guessed rule reports. Using the data from 50 simulated subjects, rule reports were randomly generated for each of 100 trials. The total set of 50 subjects was simulated 100 times in this fashion. Based on this extensive simulation, "overall mean validity of guessed rules did progressively underpredict our subjects'

proportions correct judgments (slope = 1.65, intercept = -.38), and all 100 simulations yielded slopes greater than one and intercepts less than zero" (p. 27). These data showed that producing rule reports in this task was not a purely random process.

Clearly, the conclusion that classification judgments cannot be made without verbal awareness depends on two assumptions: (a) the instrument measures what it claims to measure, and (b) the close relationship between classifications and rule validities is not preordained by the design of the experiment. If either of these two assumptions is not satisfied then the interpretation is invalid. This raises the issue of the appropriateness of the analysis used in these studies: the regression analysis on classifications and computed rule validities.

The interpretation of this study has implications that go beyond a simple critique of the growing list of artificial grammar experiments. Its conclusions touch issues not only of classification and verbal awareness, but of recognition memory, representation of knowledge, and the status of consciousness in human psychology. Dulany et al. (1984) have been widely cited as evidence that all classification behavior is verbalizable (Ericsson & Simon, 1984; Perruchet & Pacteau, 1990). To date, this argument has served as powerful evidence against the implicit knowledge hypothesis. Before rejecting the notion of implicit knowledge in artificial grammar tasks, therefore,

it may be worthwhile to examine this study again and ask if its conclusions are fully warranted.

The present study reexamined the relationship between classification and reports in one version of the artificial grammar task. Experiment 1a attempted to replicate the findings of Dulany et al. (1984). Experiments 1a and 1b describe the outcome of several computer simulations and consider them in terms of the discrepancy found between the simulations reported by Reber et al. (1985) and Dulany et al. (1985). A model was developed for generating reports in this task and is discussed in the context of reinterpreting the results reported by Dulany et al. (1984, 1985).

The purpose of this experiment was twofold: first, to attempt to replicate, as nearly possible, the important experimental results and simulation results of Dulany et al. (1984, 1985) and second, to provide a set of data upon which a model of featural rule reporting in this task could later be developed. It was hypothesized that the methods of Dulany et al. were sound and that the results were replicable. It was expected that subjects would be able to correctly classify test items at a rate equal to or approaching those reported by Dulany et al. (1984). Likewise, the instruction manipulation (memory versus rule-discovery instructions) should follow Dulany et al. (1984) and have no effect on classification judgments and no effect in the relationship between classification performance and the mean validities of featural rules. It was also expected

that the lengths of featural rule reports by subjects would follow the same as those in Dulany et al., in which the reports of items judged grammatical were longer than items judged nongrammatical, and increased with item length.

Two simulations of guessed rule reporting were conducted to check the accuracy of the algorithm and the methods used in Dulany et al. (1985). The first simulation used a set of 50 simulated subjects in which classification judgments were generated by the experimenter.

In pilot testing this experiment it was found that human subjects regarded some of the items in the test sets as compellingly grammatical or nongrammatical; that is, they classified some items the same way at a rate of 80% or higher. The use of "simulated" subjects, in which correct judgments for each string are generated by a simple probability function, fails to account for this phenomenon. To provide a contrast with the simulated subject group, the second simulation used the actual judgment data generated by the human subjects who participated in the artificial grammar experiment.

Both simulations of guessed reporting were compared to the results published by Dulany et al. (1985) with respect to the slope and intercept of the regression analysis and to the correlation between judgments and rule validities. It was hypothesized, following Dulany et al., that mean rule validities would progressively underpredict percent correct classification judgments.

Method

Subjects

The subjects were 43 undergraduates at the University of Florida who participated as part of a requirement for an introductory psychology class. The data from 10 other subjects were discarded: seven for marking single items as both grammatical and nongrammatical, one for making no judgment on over 10 items, and two for inattention during the study phase. Subjects were tested in groups of 8 to 13; 22 participated in the "memory" condition, and 21 in the "rule discovery" condition.

Procedure

The procedure for the experiment follows closely that conducted by Dulany et al. (1984). The finite-state grammar used to generate the grammatical letter strings is seen in Figure 1. The letter strings used during the study and test phases are shown in Appendix B. Nongrammatical strings were created by including invalid letters or reordering letters within grammatical strings; nongrammatical strings are underlined at the point of violation. During the study phase, subjects viewed for 10 minutes 20 grammatical letter strings projected on a screen using an overhead projector. The entire set of 20 grammatical items was viewed simultaneously. This corresponded to the "all" condition in Dulany et al. (1984) in which subjects saw the entire set of grammatical study items on one screen.[4]

Subjects in the memory condition were instructed as follows:

> This is a simple memory experiment. You will see items made of the letters M, R, T, V, and X. The items will run from three to six letters in length. You will see a set of 20 items. Your task is to learn and remember as much as possible about all 20 items. (Dulany et al., 1984, p. 544)

To subjects in the rule discovery condition, it was stressed that:

> The order of letters in each item of the set you are about to see is determined by a rather complex set of rules. The rules allow only certain letters to follow other letters. Since the task involves memorization of a large number of complex strings of letters, it will be to your advantage if you can figure out what the rules are, which letters may follow other letters, and which ones may not. Such knowledge will certainly help you in learning and remembering the items. (Dulany et al., p. 544)

The test phase consisted of 100 letter strings printed on four pages. The 25 grammatical and 25 nongrammatical letter strings were randomly ordered and printed on the first two pages; the same letter strings were reordered and printed on the next two pages. All subjects were instructed in the test phase as follows:

> The order of letters in each item of the set you saw was determined by a rather complex set of rules. The rules allow only certain letters to follow other letters. About half of the lines of letters on your answer sheets follow the rules used to generate the lines of letters you saw earlier, and about half of them violate the rules in some way. For every item, you have to do two things. One, you have to decide whether or not the line of letters follows the rules, and then you have to justify your decision.

The procedure for making a judgment was demonstrated by the experimenter. The letters KZK were written on a chalkboard.

The experimenter underlined or crossed out a portion of the sample item to illustrate the procedure for making a response. It was stressed that each item should contain only one type of mark, and that a mark should be made for each item.

Simulation of Featural Rule Reporting

There were two simulation runs performed. For the first run, 50 simulated subjects were created by generating 100 classification judgments for each subject (recall that there were 50 letter strings in the test set, and each was judged twice). The distribution of percent correct judgments for the group was created by estimating the percent correct for each of the 50 experimental subjects reported in Dulany et al. (1984, p. 547). The percent correct judgments (*M* = 64.4%, *SD* = 7.0%) of these 50 simulated subjects followed closely the classification performance of the subjects in Dulany et al.

For the second run, the classification data from the 43 human subjects were used. For both simulation runs, the procedure was the same. For each of 100 trials, for each subject, the program randomly generated a report length ranging from 2 to the number of letters in the letter string, then randomly generated the starting position of the report in the string, constrained only by the length of the report. Thus, there were 100 rule reports associated with 100 letter strings for each subject in both the "simulated" subjects group and the "real" subjects group. The program

was coded from the random-guessing algorithm published by Dulany and Carlson (1985).[5]

Results

Judgments and Conditions

The mean percentage of correct classification judgments for the 43 subjects in the present study was 63.4%, compared to 64.9% for the experimental group in Dulany et al. (1984). The rule discovery group (M = 66.8%, SD = 8.1%) scored higher than the memory group (M = 60.3%, SD = 8.7%), $F(1, 41) = 6.67$, $p < .05$. Most of the subjects had better-than-chance performance as indicated by a score of 60% or higher: 18 out of 21 subjects in the rule-discovery group and 11 out of 22 in the memory group.

Table 1 shows the mean lengths of featural rule reports calculated for subjects' grammatical and nongrammatical responses. The data were analyzed in a 2 x 2 x 2 (Instruction x Response x String length) ANOVA. As expected, subjects marked more letters per report for their grammatical responses than nongrammatical responses, $F(1, 41) = 47.0$. The length of reports also increased with String length, $F(1, 41) = 84.4$. Response interacted with String length, $F(1, 41) = 14.4$. All effects were $p < .0001$.

Table 1

Mean number of letters reported per featural rule as a function of response and string length.

	String length			
Response	3	4	5	6
	Memory group			
Grammatical	2.7	3.0	3.4	4.0
Nongrammatical	2.2	2.3	2.7	2.9
	Rule discovery group			
Grammatical	2.7	3.2	3.6	4.3
Nongrammatical	1.9	1.9	2.1	2.3
	Combined			
Grammatical	2.7	3.1	3.5	4.1
Nongrammatical	2.0	2.1	2.4	2.6

There was no effect of Instruction on report lengths, $F(1, 41) < 1.0$, but the analysis revealed an Instruction x Response interaction, $F(1, 41) = 4.47$, $p < .05$. Except for this interaction, and the unexpected effect of Instruction on judgments, the results appeared to be comparable to Dulany et al. (1984).

Rules and Judgments

Figure 2 displays the scatter plot of calculated mean rule validities and percent correct judgments. Using the G3 computation, mean rule validity was .647 (the details for computing mean rule validities are in Appendix A). The regression equation yielded a slope of 1.09 and an intercept of -0.07, $r = .95$. The regression line was not significantly different from the unit slope, $t(41) = 1.62$, $p > .10$.

Simulation of featural rule reporting

For the "simulated" subject group, the overall mean rule validity was .60, as compared to 64.4% correct judgments. Mean validity underpredicted mean percent correct judgments (slope = 1.24, intercept = -0.1). The regression line was significantly greater than the unit slope, $t(48) = 2.29$, $p < .05$. Thus, the nature of the test set of items was such that randomly generated item fragments had a high (approximately .60) level of predictive validity. However, randomly generated items still underpredicted percent correct judgments. Notably, the correlation was $r = .86$, which was as high as the correlation reported for the experimental groups in Dulany et al. (1984).

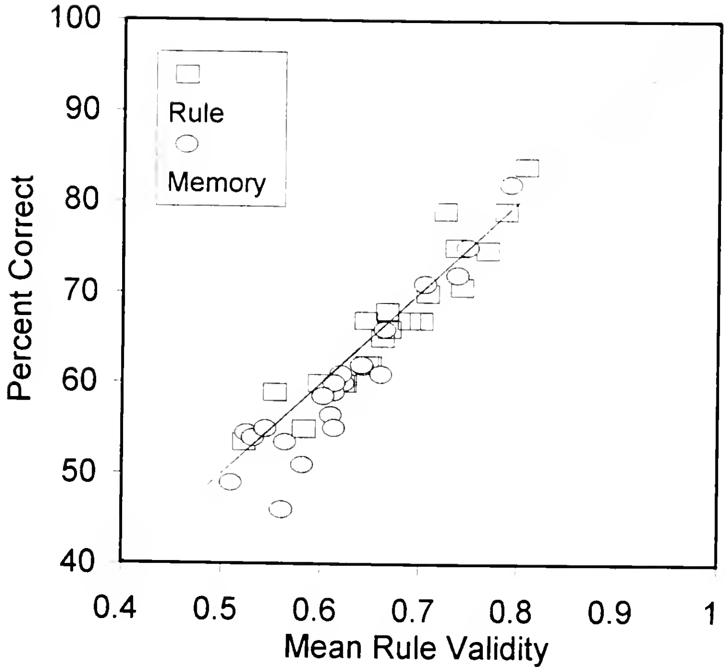

Figure 2. Scatter plot of percent correct judgments and mean rule validities of "real" rule reports for each of 43 subjects in Experiment 1a.

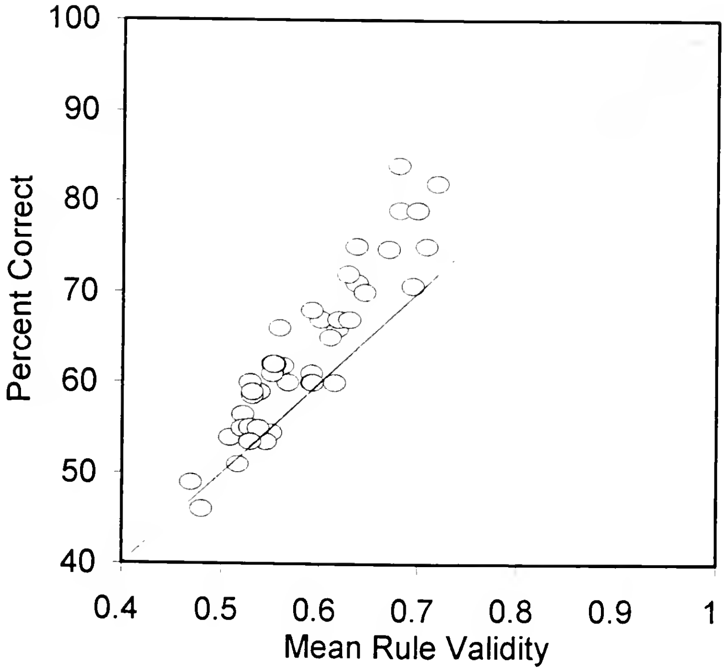

Figure 3. Scatter plot of percent correct judgments and mean rule validities of randomly generated rule reports for each of 43 subjects in Experiment 1a.

For the "real" subject group, in which random featural reports were generated for each of 43 real subjects' judgments, mean rule validity was .586, compared to 63.4% judgments. Mean validity underpredicted mean percent correct judgments (slope = 1.3, intercept = -0.12). The intercept of the regression line was significantly less than zero, $t(41) = -2.54$, $p < .05$; the slope was greater than the unit slope, $t(41) = 3.55$, $p < .001$. The correlation (r = .92) was higher than that reported by Dulany et al. (1984). See Figure 3.

Discussion

The experiment replicated in large measure the experiment carried out by Dulany et al. (1984). The results closely followed, in terms of length of reported featural rules and the relation of mean rule validities and percent correct. The only unexpected finding was that the rule-discovery group outperformed the memory group in the classification task. Previous research had shown that subjects instructed simply to memorize letter strings during the learning phase sometimes scored higher on the classification task than those informed of the rules underlying the letter strings (Reber, 1976; Reber & Allen, 1978). It has been pointed out that this particular version of the artificial grammar task is somewhat different than previous versions of the task (Reber et al., 1985). However, a number of studies have failed to replicate the original "instructional set" effect for classification

judgments; the instruction manipulation seems rather weak (Dienes, Broadbent, & Berry, 1991; Millward, 1980; Reber et al., 1980, Experiment 2).

Both of the simulations conducted also supported the results of the simulation reported by Dulany et al. (1985). Purely random guessing of rules did not account for the relationship between judgments and reported rules, at least for subjects who learned from having viewed the letter strings during the study phase. The methods of Dulany et al. (1984, 1985) appear to be sound, and the basic findings replicable.

As previously stated, however, the aim of this study was not to question the data presented by Dulany et al., but to question their interpretation of performance of the task. Therefore, it is worthwhile to review some of the alternative interpretations offered to explain the process by which subjects' decisions are made in making their classification judgments.

It has been argued that rule reports are measuring verbal awareness in the task, and that this verbal awareness leads causally to classification judgments: judgments arise from consciously identified rules that are represented by the featural rules reported for each judgment (Dulany et al., 1984). In support of this position, Dulany et al. considered but rejected three alternative explanations:

1. The featural rules were guessed after the classification judgments were made. As independent computer

simulations have shown, pure guessing cannot account for the relationship between mean rule validity and percent correct. It seems unlikely, in any case, that subjects would be unable to recognize frequently occurring letter bigrams and trigrams. Indeed, subjects frequently are able to recall salient bigrams and trigrams in post-experimental interviews (Reber & Allen, 1978).

2. The featural rules emerged from an unconscious representation of the stimuli, a "conscious reconstruction of some aspect of a nonconscious grammar" (Dulany et al., 1984, p. 553). This was suggested as a possible interpretation, but its likelihood not explored, nor was any support for the position cited.

3. Classification judgments and reported featural rules exist as independent knowledge. This is essentially the position of Reber and others who support an abstractive view of the knowledge acquired in this task (Reber, 1976; Reber et al., 1985). "The reported rules are learned in parallel with the unconscious grammar and then are recalled as cued by the string at hand" (p. 553). The first problem with this account, according to Dulany et al., is the absence of "linking assumptions, which is to say a description of a process that would strongly relate assumed amounts of unconscious grammatical learning to the observed mean validities of reports" (p. 553). A second problem invokes an appeal to parsimony: it is easier to explain the effect of validities and correct judgments as one of conscious

control, rather than as separate systems apparently doing the same thing.

The relationship of classification and recognition measures figures prominently in larger issues of concept formation, as in the debate over prototype versus exemplar theories. Dissociations between classification and recognition are cited as support for prototype theories, a variety of a separate-systems theory (e.g., Hayes-Roth & Hayes-Roth, 1977). However, the same data cited in support of prototype theories were found to be consistent with single-system exemplar models, supposing that classification and recognition were computed separately (Nosofsky, 1988). It was suggested that the differences in classification and recognition measure found in some tasks were the result of different "decision rules" used in each process. In this view, classification involved a comparison between a target category and a contrast category, whereas recognition may be determined by overall summed similarity of a probe to all exemplars stored in memory. "Thus, classification and recognition may often be based on common representational substrates, but different decision rules may underlie performance in each task." (Nosofsky, 1988, p. 707). This analysis supports a single-system exemplar view of classification behavior as opposed to a separate-systems view (e.g., Reber, 1989), but preserves the notion that classification and recognition are computed separately.

The reporting technique used in this experiment may more properly be thought of as one of recognition, rather than of verbal awareness (Broadbent, 1991; Reber et al., 1985). In the analysis of the task by Dulany et al., the featural rules supplied by subjects during the task tell us nothing about how the particular item fragment was selected by the subject, whether other fragments were considered, or if other item fragments had an effect on the subject's judgment. However, if we work from the assumption that item fragments are based on recognition memory, it is possible to begin to explore schemes for computing the familiarity of item fragments. Under this assumption, the reporting technique touted by Dulany et al. as providing a reason for a classification judgment is seen simply as a cue for the subject to supply a featural rule based on an item fragment's perceived familiarity.

The issue of classification versus recognition in this task must take into account the instrument used in this version of the artificial grammar task: the relationship between classification and mean rule validities as revealed by the regression analysis. The next experiment turns to the problem of postulated processes, and the "linking assumptions" that cause classification judgments and rule validities to follow each other over a large group of subjects. Put more succinctly, "Is there some more systematic way that rules might come to track judgments without controlling them?" (Dulany et al., 1984, p. 553).

Notes

[1]In this grammar, each circle (S_0, S_1, . . ., S_5) denotes a node and each arrowed line denotes a directed transition between nodes. Grammatical letter strings are generated by following a path through the grammar from node to node from beginning to end. At each transition between nodes a symbol (M, T, V, X, or R) is generated. For example, the path represented by S_0, S_1, S_3, S_5 produces the grammatical letter string MVT. The letter string VXX is nongrammatical: There is no valid path through the grammar that can generate that sequence of letters.

[2]Simply put, the validity of a featural rule is a proportion based on the number of times the featural rule appears in items of a selected category and the number of times it appears in all items in all categories. For example, if MT appears in x grammatical items and y nongrammatical items then the computed rule validity for this featural rule is x / (x + y). See Appendix A for further information.

[3]Unfortunately, the details of this simulation are no longer available (Arthur Reber, personal communication).

[4]The "all" display during study was contrasted with a display in which study items were viewed individually; there were no differences found between the "all" and "individual" groups (Dulany et al., 1984). Thus, there was no particular benefit to seeing all the study items at the same time in this particular version of the artificial grammar task.

[5]The algorithms were in fact slightly different. Dulany and Carlson's algorithm estimated the lengths of the featural reports published in Dulany et al. (1984). In the present study, the report lengths were generated completely at random.

CHAPTER 3

EXPERIMENT 1B

For this experiment a model was developed of item fragment recognition and featural rule report in the artificial grammar task. The model was instantiated in a computer program that was written to simulate the process of producing reports in this task. The outcome of the simulation was evaluated in terms of the model's psychological validity.

The model of producing rule reports in the artificial grammar task began with the assumption that classifications and reported featural rules are computed independently. It also was assumed that rule reports involve recognition processes rather than verbal awareness for the contingencies behind classification judgments. Following Nosofsky (1988), the classification of objects involves the comparison of objects judged to be within the target category (items judged grammatical) and outside the target category (items judged nongrammatical). The model developed in this experiment did not deal explicitly with the processes involved in making judgments; there already exist a number of plausible computational accounts of how classifications are made in artificial grammar tasks, both symbolic (Servan-

Schreiber & Anderson, 1990) and connectionist (Dienes, 1992).

How are the reports produced? After making judgments, subjects in Experiment 1a were required to justify their decisions by marking some portion of a test item. The data from Experiment 1a and from Dulany et al. (1984) followed some reliable trends. Reports of items judged grammatical were longer than items judged nongrammatical and increased with the length of the test item. Reports of items judged nongrammatical averaged about two in length. In this model of rule reporting, subjects searched items they judged grammatical for item fragments that appeared familiar; items that were judged nongrammatical were searched for a pair of unfamiliar letters. The model also followed the instructions given to the subjects in the test phase of the experiment: First, to make a judgment about an item, then provide a rationale for having made that judgment.

An important feature of the model is the assumption that subjects did not search items exhaustively and often failed to find an optimal report. That is, a few item fragments were randomly considered, and the best (the most familiar or least familiar) item fragment reported. Previous research had shown that subjects were fairly accurate at discriminating between letter bigrams that appeared in a study phase and new, or unfamiliar, bigrams (Perruchet & Pacteau, 1990). The same item fragment might be considered several times; most fragments in an item were

not considered. Furthermore, it was assumed only that subjects who scored better than chance in classification judgments had any memory of the study items; subjects who scored less than chance were assumed to be responding randomly in producing featural rule reports.

In this model of rule reporting, subjects first classified an item, then briefly searched the item for evidence of its grammaticality or nongrammaticality. Items judged grammatical were searched for item fragments that appeared familiar; nongrammatical items were searched for pairs of letters that appear unfamiliar. The familiarity of an item fragment was based on its summed frequency of appearance in the set of study items (Nosofsky, 1988). It is reasonable that subjects should be most familiar with the items, and therefore the item fragments, that they were instructed to remember: The items that were viewed for 10 minutes during the study phase of the experiment. The model incorporated this assumption by computing the familiarity of an item fragment viewed during the *test* phase against its number of occurrences in the *study* phase. The best item fragment was reported as a featural rule.

A computer program was written to simulate this model of featural rule reporting, and was run against the judgment data from Experiment 1a. It was hypothesized that the mean rule validities of featural rules produced by this model would predict percent correct classification judgments as

accurately as the reports produced by the subjects in Experiment 1a.

A fully explicit, exemplar account of classification and rule reporting in this task starts with the selection of an item fragment followed by the judgment of the item (Dulany et al., 1984). In this view, the close association between the percentage correct judgments and mean rule validities supports that model. If, however, the same association could be demonstrated based on separate computations and separate sources for classification and rule reports, this would suggest the possibility that separate processes were used in performance of the task.

Method

The simulation uses the judgment data collected from the 43 experimental subjects in Experiment 1a. For each grammatical or nongrammatical judgment, item fragments are generated and considered, and a featural report is produced. Each featural rule can be characterized by a rule length (from 2 to the maximum number of letters in the item) and rule position (which specifies the position of the first letter in the rule).

The familiarity of a item fragment was calculated based on its frequency in the set of study letter strings. The item fragment is counted from the rule length and rule position of all strings in the study set, regardless of the length of the string. This corresponds to the G2 method of deciding whether a given featural rule occurs in an item,

and is less constrained than the G3 calculation (Dulany et al., 1984).[1]

The individual percents correct for each subject were calculated. Twenty-nine subjects demonstrated their having learned to discriminate between grammatical and nongrammatical test items by scoring 60% correct or higher during the test phase, a rate better than chance. Fourteen subjects scored less than 60%, and were assumed not to have learned to discriminate test items. These 14 subjects were assumed to simply be responding randomly, both on their classification judgments and their rule reports.

For the 14 subjects who scored less than 60% correct judgments, the simulation did not consider the familiarity of item fragments, but simply reported the first item fragment generated. This corresponds to random responding of rule reporting. For the 29 subjects who scored 60% or more, the following algorithm was used:

```
IF item is Judged Grammatical THEN
DO 7 TIMES
        Randomly generate an item fragment
        Compute the familiarity of the item fragment
                IF familiarity > top familiarity score
                        Save item fragment
                        Save top familiarity score
LOOP
IF item is Judged Nongrammatical THEN
DO 7 TIMES
```

```
        Randomly generate an item fragment
        Compute the familiarity of the item fragment
            IF familiarity < least familiarity score
                Save item fragment
                Save least familiarity score
            IF familiarity = 0 then EXIT LOOP
    LOOP
    Report the saved item fragment as featural rule
```

Following the model of rule reporting in this task, reports generated on grammatical trials were between two letters long and the maximum length of the item. Reports generated on nongrammatical trials were two letters long.

The number of item fragments available in each item is given by the formula $(n * (n + 1)) / 2$, where n is the number of letters in the item. Thus, an item 5 letters long has $(5 * (5 + 1)) / 2$, or 15 item fragments. Pilot testing of the simulation showed that 7 iterations of the loop, although it does not generate and consider all item fragments in every item, provided a close approximation between computed mean rule validity and percent correct judgments.

Results

Figure 4 displays the scatter plot of mean rule validities and percent correct judgments. Using the G3 computation, mean rule validity was .62. The regression analysis yielded a slope of 1.04 and an intercept of -.01, $r = .94$. The intercept was not significantly different from

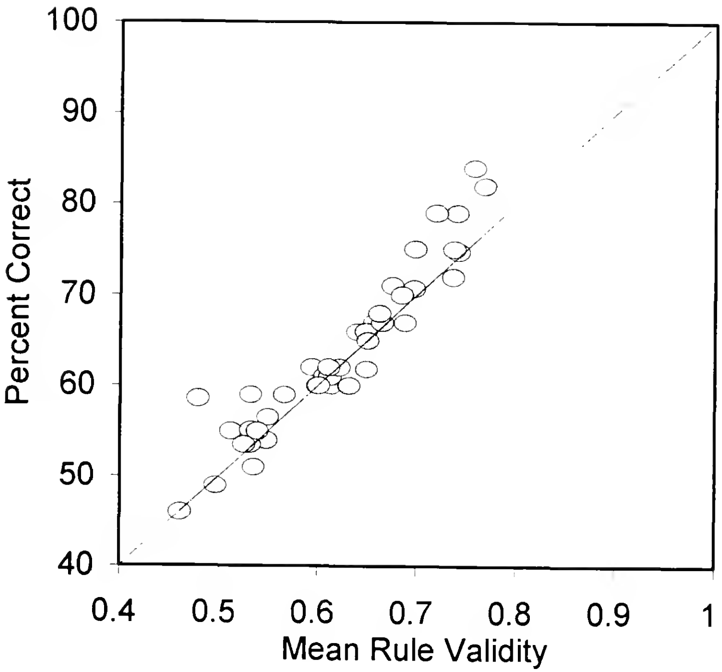

Figure 4. Scatter plot of percent correct judgments and mean rule validities of simulated "familiar" rule reports for each of 43 subjects.

zero, t(41) = -0.25; the regression line was not significantly different from the unit slope, $t(41) = 0.65$.

The mean lengths of featural rule reports were calculated (see Table 2). As specified in the algorithm, all nongrammatical reports were two letters long. Grammatical reports increased with item length, as found in Experiment 1a.

Table 2

Mean number of letters reported per featural rule as a function of response and string length.

	String length			
Response	3	4	5	6
Grammatical	2.3	2.5	2.9	3.1
Nongrammatical	2.0	2.0	2.0	2.0

Discussion

In assessing the psychological validity of the algorithm for recognizing item fragments and producing reports, it is useful to remember what the algorithm did not do. It did not search exhaustively for familiar or unfamiliar item fragments. The search was undirected; no attempt was made to keep track of item fragments which were

generated and considered. The algorithm did not account for item length in assessing the frequency of fragments; fragments were counted only from the beginning of all items. The algorithm was not dynamic in that items viewed during the test phase were not remembered, and had no effect on the reported featural rules.

Most importantly, there was no attempt to "scale" the efficiency of the algorithm within the two sets of subjects. Subjects who were able to discriminate grammatical from nongrammatical items at a rate higher than chance ranged from 60% correct to 85% correct. It might plausibly be suggested that the subjects at the top end of the scale learned more of the "correlated grammars" needed to classify items than those scoring only 60% and might also be more efficient at recognizing and reporting rules, but the algorithm made no allowance for this. From an examination of the scatter plot of Figure 4, there appears to be a slight tendency for mean rules validities to underpredict percent correct judgments for a few of the highest scoring subjects. However, the regression analysis was not sensitive enough to detect any "progressive" underprediction. Despite these many limitations, the simulation was still able to produce featural rules which closely tracked the judgments made by the experimental subjects.

Asking subjects to report familiar and unfamiliar portions of items provides clues to the information that

subjects are using as they are making judgments, but selection and report of the featural rules themselves, the basis upon which the present analysis rests, incompletely specify what the subject knows about an item. As found in Experiment 1a, Dulany et al. (1984) found that many subjects--despite the experimenter's careful instructions--insisted on reporting more than one item fragment. Subjects also tended to mark part of an item "grammatical" and another part "nongrammatical." This could be interpreted as the subject's way of telling the experimenter "I believe this item to be both well-formed and not well-formed." A more likely explanation would be that the subject first classified the item, but found a portion of the item familiar and a portion unfamiliar, and reported each.

As an instrument for detecting dissociations between classifications and recognition, the regression analysis reveals little more than the fact that reports are not made at random. The correlation coefficient is equally uninformative. It was noted that correlations between randomly generated rules and percent correct were as high or higher than those reported for real subjects; why would the correlation between rule validity and judgments decrease as subjects learn the task and produce "nonrandom" reports? Dulany explains this as nothing more than sampling error. If this were the case then real data would reliably show correlations higher than those for randomly generated data.

If it is assumed that computations were made independently for classification and recognition of item fragments (upon which the featural rule reports were based), then it suggests one possible interpretation for the relatively low correlation coefficient found among subjects who participated in the study phase of the experiments. Data points above the unit slope, in which mean rule validity underpredicts percent correct judgments, represent good classification performance and poorer explicit recognition of item fragments. Points below the unit slope, where mean rule validity overpredicts percent correct, represent good recognition, and poorer classification performance based perhaps on poor integration. To the extent that there are differences between classification and explicit recognition of the test items, some subjects may tend to demonstrate superior classification performance of test items, some will show superior recognition of study items, and the correlation between the two measures will go down.

The main issue deals with the question of the instrument's sensitivity to the hypothesized difference between classification and reported rules. The magnitude of the difference between randomly generated rule reports and those provided by experimental subjects is quite small. That is why a crude recognition algorithm, such as that used in this model, can "cover the distance" between randomly generated reports and subjects' report data: there's very

little distance to cover. Although the instrument may, with modification, be useful in detecting dissociations between classification and explicit recognition, the presence or absence of the effect is nondiagnostic for the issue over classifications and verbal awareness in the artificial grammar task. Given this particular task, where the stimuli are arranged such that explicit recognition and classification follow each other closely, it is difficult to find a set of conditions where the outcome is not preordained.

Is there some way that featural rules might come to track judgments without controlling them? It has been shown that, in this particular task, it is hard to come up with a way of producing reports (short of random guessing) that will *not* track judgments. The fact that featural rules track judgments says as much about the stimuli as the phenomenon under investigation.

Classification and Recognition

The instrument used in this study does not allow one to definitively distinguish between two competing hypotheses: one in which a rule is consciously considered and leads to classification, or one in which classifications and reported rules are computed separately (Dulany et al., 1984). The aim of this study was relate the artificial grammar experiments to larger issues of concept formation and recognition, and to demonstrate that the hypothesis of

separate computations is (at least) as likely as one of fully conscious control.

In subjects with unimpaired memories, measures for classification and explicit recognition are inevitably confounded (Turner & Fischler, 1993). Evidence for separate computation of classification and recognition has come from work with amnesic patients. In one study involving artificial grammars, amnesic subjects performed as well as normals in a classification task, but were impaired on a test of letter string recognition (Knowlton, Ramus, & Squire, 1992). Simply put, classification and recognition involved separate, but co-occurring, computations.

The Role of Simulation

The simulation reported in Experiment 1b did not "prove" the hypothesis of separate computations in the artificial grammar task. Rather, it demonstrated the feasibility of separate computations. Rule reports generated by a simple algorithm on one set of items (the study items) were matched to judgments made by subjects on a separate set of items (the test items). The reporting measure, computed independently of the judgments, followed the judgments as closely as any produced by real subjects.

As a tool for research, the simulation will allow one to generate testable hypotheses and predictions for performance in experimental follow-ups. The model predicts, for example, that subjects would be able to take test scoresheets on which are printed randomly generated

classifications and mark familiar and unfamiliar item fragments for each, with the same results as described in Experiment 1b.

Importantly, the simulations in both experiments gave insight into the structure of the instrument that is not readily apparent. The random reports generated in Experiment 1a and in Dulany et al. (1985), when compared with real subjects' data, showed clearly how close the "ceiling" and the "floor" are in this task. Simulating the experiment also suggests ways in which this instrument can be modified to detect dissociations between classification and explicit recognition.

Note

[1]For a discussion of the relative merits of the G2 versus the G3 calculations for rule validity, see Appendix A.

CHAPTER 4

EXPERIMENT 2

In building a case for the independence of implicit and explicit knowledge, Reber (1976) cited the fact that subjects in his experiments were unable to explain the basis for their classification judgments. It was claimed that memory-instructed subjects made judgments in an intuitive, "wholistic" fashion, implying that the rule-discovery subjects made their judgments consciously and deliberately. However, no data on verbal reports were presented to indicate that these rule-discovery subjects, presumably possessed of a base of knowledge that was verbally accessible, were making verbalizable judgments. In fact, little data have been presented in the artificial grammar tasks that demonstrates different modes of processing based on intention to learn (Brody, 1989).

More importantly, there has been little evidence from the artificial grammar experiments for difference levels of verbal behavior based on some variable, such as intention to learn. To the extent that verbal knowledge does not accurately predict one's performance on a given task, there is said to be independence, or "dissociation" between explicit knowledge and implicit, tacit knowledge. Such dissociations have been found in experiments involving the

control of so-called complex systems (Berry & Broadbent, 1984, 1986; Hayes & Broadbent, 1988). However, the effect is subtle, and demonstrating dissociations based on verbal report in these complex system tasks is the subject of much dispute (Dulany, 1991; Sanderson, 1989).

The case for independent processes depends on two points: differences between groups in verbal report and differences in task performance. Berry and Broadbent (1984) examined the relationship between performance on a cognitive task and the explicit, or reportable, knowledge associated with that performance. Subjects controlled a complex computer-simulated system by supplying input in the form of integers and observing the resulting output. The output was based on the following formula:

$$O = ((I \times 2) - 01 + R)$$

where o = current output, I = input, 01 = the previous output, and R = either -1, 0, or 1, randomly selected.[1] Thus, there was no one-to-one mapping from input to output. The subjects' task was to achieve and maintain a given output goal. Questionnaires were administered after the task was completed in order to assess subjects' verbal knowledge of the behavior of the system, and the rules they had developed in order to control the system.

In one experiment, practice with the system improved subjects' ability to control the system, but had no effect on the ability to answer related questions. In a second experiment, one group of subjects was given verbal training

prior to the start of the task; a second group simply performed the task. This verbal training improved subjects' ability to answer questions but had no effect on control performance. In a third experiment, verbal instruction combined with concurrent verbalization led to an improvement in control scores. Verbalization alone, however, had no effect on task performance or question answering. Results such as these show the independent, but related, nature of implicitly acquired knowledge and verbal (or declarative) knowledge. "These data allow for stronger conclusions than are possible on the basis of a simple demonstration that subjects cannot give an adequate verbal account of their overt behaviour" (Hayes & Broadbent, 1988, p. 250).

However, much of the work done with artificial grammars does rely on "simple demonstrations" of subjects' inability to verbally justify the reasons for their decisions or solutions to tasks. One of the reasons that Dulany's criticism is so important is the fact that no one on the implicit learning "side" has been able to show dissociations between performance and verbal abilities in groups of subjects based on a variable, such as type of study or intention to learn, differences that should exist according to the implicit learning hypothesis.

Mathews, Buss, Stanley, Blanchard-Fields, Cho, and Druhan (1989) attempted to dissociate performance and verbal report in an artificial grammar task. During an initial learning phase, subjects received memory or rule-discovery

instructions. During the test phase, in what the researchers called a "teach-aloud" procedure, subjects provided concurrent verbal descriptions of the rules on which they based their grammaticality judgments. During breaks between blocks of trials in the test phase, memory subjects also were asked to recall grammatical strings, or string fragments, from the study and test phases.

A third and fourth group of subjects, the "yoked" groups, did not participate in an experimental study phase, but made their judgments during a test phase using transcripts of the memory and rule-discovery subjects' concurrent verbal protocols. The yoked subjects' performance on the classification task was better than that of a control group, but was worse than that of the two experimental groups. The results showed that the experimental subjects had some verbal knowledge of the rules used to order the letter strings, as evidenced by their ability to relay rules to the yoked groups, but that their verbal knowledge was incomplete.

Notably, they did not find reporting differences between the memory and rule-discovery groups, nor performance differences between the two yoked groups. If the rule-discovery subjects had better verbal awareness of the rules by which they made their judgments, this should have permitted better verbal report on the "teach aloud" task, and better performance by the subjects yoked to the rule-discovery group. Thus, Mathews et al. were able to

show that verbal ability lags behind classification performance in the artificial grammar tasks, but were not able to fully dissociate performance and verbal awareness based on intention to learn, as the Broadbent studies had been able to do. This failure to dissociate judgments and verbal measures of awareness are a weak point of the implicit learning research.

The present experiment attempted to find dissociations between judgments and two measures of memory for study items: one of recall and one of recognition. In addition to a recall test of item fragments after the study phase, the experiment introduced a modified version of the test phase reporting task from Dulany et al. (1984). The experiment also included a group of subjects who learned the task as a set of logical rules. This "rule-training" for logical rules results in substantially higher performance in unspeeded test trials, and allows a direct comparison of rule-discovery and fully explicit learning (Turner & Fischler, 1993).

The first measure, the recall test, was administered in the form of a questionnaire at the end of the study phase. It was hypothesized that rule-discovery subjects, informed of the nature of the task and fact that the study items were rule-ordered, would be better able to answer general questions about the study items than the memory group.

The second measure, the recognition test, was a modified version of the reporting technique used by Dulany

et al. (1984). In the present Experiment 1a, it was argued that the stimuli and design of the experiment were such that there could be little alternative to the close association between percent correct judgments and computed mean rule validities. Because of this, the design of the reporting measure is inadequate for showing differences between groups in the ability to recognize and report featural rules. It is possible, however, that the reporting technique could be altered in such a way as to permit different levels of performance of the task.

In particular, there are two features of the technique that insured a high correlation between percent correct and mean rule validity and the resulting regression line. First, nearly every portion of every letter string has predictive validity in the sense that it occurs more frequently in one grammatical category than the other. Second, there are no constraints on the length of allowable reported item fragments; the longer the reported item fragment the better the item's predictive validity. Forcing subjects to be more specific about their reports by constraining their reports to one or two letters would introduce more variability into the rule reports and allow differences in the level of predictive validity of rule validities. It was hypothesized that rule-discovery subjects would be better able to recognize item features that determine the items nongrammaticality.

It was hypothesized that the rule-training group, having had specific training, would show near-perfect knowledge of the rules underlying the production of the letter strings; the regression analysis, however, would not predict performance as accurately as for the two observation groups. The comparison of the rule-training group and the rule-discovery group might also shed light on theories of explicit processes.

Method

Subjects

Sixty-nine undergraduate college students served as subjects. The data from three subjects was dropped for failure to follow instructions. All were voluntary participants satisfying a requirement for an Introductory psychology course.

Materials

The materials followed those used by Reber et al. (1980) and Turner and Fischler (1993). A finite-state grammar represents the rules employed in the study to generate the grammatical letter strings; see Figure 5. Twenty-one letter strings were selected for use in the study phase. Twenty nongrammatical strings were constructed such that they each resembled a grammatical string, but contained one violation of the five logical rules needed to correctly discriminate grammatical from nongrammatical strings.

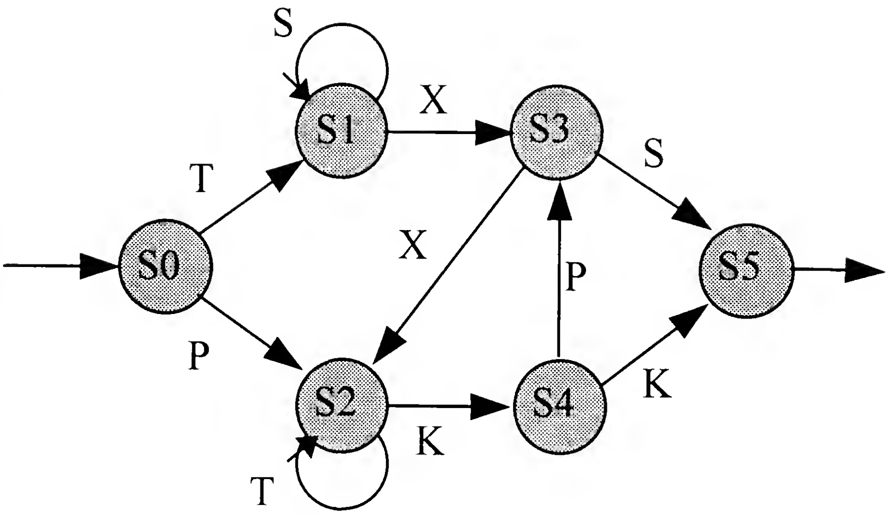

Figure 5. A representation of the finite-state grammar used to generate the grammatical letter strings in Experiment 2.

Generating nongrammatical items in this way permitted the creation of a small set of logical rules that could be easily memorized by subjects and used to fully discriminate grammatical from nongrammatical items. The logical rules, and their associated Rule Numbers, were:

(1) The first letter in a string must be T or P
(2) The last letter must be S or K
(3) The letter pair "P P" cannot appear in a string
(4) If the first letter is T then the next letter must be X or S.
(5) If the first letter is P then the next letter must be T or K.

These rules were printed on sheets of paper, in the order described above, for use by the rule-training groups. All letter strings are shown in Appendix C; nongrammatical strings are underlined at the point of violation. The stimuli were presented and all responses taken on IBM-compatible PC microcomputers.

<u>Design and Procedure</u>

In the study phase, subjects were assigned to one of three groups based on type of Training (memory, rule-discovery, or rule-training). There were 22 subjects in each group. In the memory condition, subjects were given the same instructions as the memory group in Experiment 1a. For the rule-discovery group, the subjects were read the following:

> The order of letters in each item of the set you are about to see is determined by a rather complex set of rules. The rules refer to which letter or letters is allowed to appear at the beginning or end of a line, and which letters can appear next to one another. Your task is to figure out the rules by looking at the lines

> of letters. After you have looked at the lines of letters, I'm going to ask you to name the rules.

In the rule-training condition, subjects learned the set of logical rules necessary to make correct grammaticality judgments.

For the two observation groups (memory and rule-discovery), three grammatical strings were displayed in each of seven frames. Each frame was displayed on the screen for 30 seconds after which the screen went blank. The set of seven frames was presented again, for a total viewing time of 7 minutes. Letter strings of similar appearance were distributed across frames, to avoid introducing structure that would make the grammar "salient." For the rule-training group, the learning phase was presented as a test of memory for logical rules. Subjects were given a sheet of paper with the rules printed on it and asked to study the rules. After studying the rules, they were verbally quizzed on the rules until they could repeat them perfectly.

After the study phase all subjects received a written questionnaire to assess their explicit knowledge of the rules. For the rule-training subjects the multiple-choice test was based on the rules they had studied during the study phase. For the subjects in the observation groups, the questions asked for memory of letters in the initial position, last position, letters that appeared as doubles, and initial letter pairs. The questionnaires appear in Appendix D.

In the test phase, the memory group learned that the letter strings were ordered by a set of rules. All subjects made classification judgments of novel letter strings presented one at a time on the computer screen. In the first half of the test phase subjects were shown 40 letter strings in random order for classification judgments. The same 40 strings were reordered and presented again.

After each classification judgment, digits appeared beneath each letter in the string corresponding to the position of the letter in the string (e.g., beneath the item "P T K P S" would appear the digits "1 2 3 4 5"). Subjects were required to enter a pair of contiguous digits to indicate the reason for their judgment. If in judging a string "grammatical," a subject could indicate that the first two letters appeared familiar or appropriate by entering the digits "1" and "2." Subjects might also enter the same digit twice, to indicate an interest in a single letter in a particular position. This is the equivalent of Dulany's paper-and-pencil procedure of underlining portions of letter strings. The computer accepted no input other than digits that appeared on the screen and were equal or contiguous to each other, thus eliminating missing or ambiguous data.

Results

Memory Group vs. Rule-Discovery Group

The first analysis compared the classification performance and reporting measures of the two observation

groups. There was no difference in mean percent correct judgments between the memory group (M = 59.7%, SD = 4.2%) and rule-discovery group (M = 60.5%, SD = 6.6%), $F(1, 42) = 0.26$, $p > .6$. The fact that the two groups scored somewhat lower than the groups in Experiment 1a could be due to the change in the stimuli used in this experiment; the strings were longer on average, and the nongrammatical strings were created by substituting a single impermissible letter in place of a letter in a grammatical letter string.

Mean Rule Validity Analysis. For subjects in the two observation groups, Figure 6 displays the scatter plot of mean rule validities and percent correct judgments. Using the "G3" computation, mean rule validity for the memory group was .577. A regression analysis yielded a line with a slope of .97 and intercept of .04, $r = .67$. The computed slope was no different than the unit slope, $t(1) = -0.14$, $p > .8$; the intercept was not significantly different from zero, $t(1) = .28$, $p > .7$.

The mean rule validity for the rule-discovery group was .585. The regression analysis produced a line with a slope of 1.09 and intercept of -0.04, $r = .74$. The slope was not significantly different from the unit slope, $t(1) = .43$, $p > .6$; the intercept was not significantly different from zero, $t(1) = -.28$, $p > .7$.

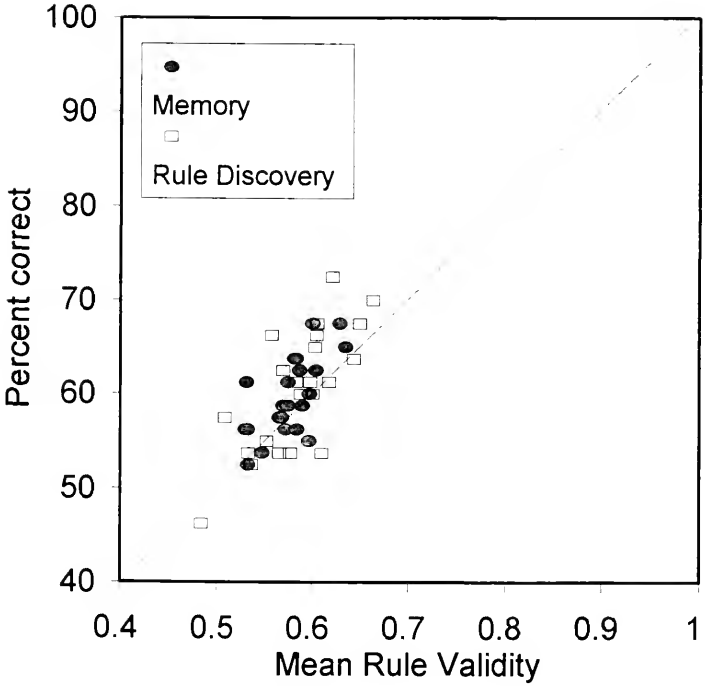

Figure 6. Scatter plot of percent correct judgments and mean rule validities for observation groups in Experiment 2.

Questionnaire Analysis. Subjects' explicit knowledge of the rules used to order the letter strings could best be described as "fair." None of the subjects in the observation groups correctly identified all 5 rules without error. Two subjects in each of the memory and rule-discovery groups correctly included all 12 letters related to the rules but incorrectly included others. In the memory group, subjects correctly identified, on average, 8.8 of the letters related to the rules and incorrectly included 5.7 letters. The rule-discovery group subjects correctly included 8.7 letters related to the rules and incorrectly included 7.2.

Observation Groups vs. Rule-Training Group

The next analysis compared the classification performance and reporting measures of the two observation groups with those of the rule-training subjects. As expected, the rule-training subjects were significantly better at the classification task (M = 85.6%, SD = 15.8%) than the observation groups (M = 60.1%, SD = 5.5%), $F(1, 64) = 93.4$, $p < .0001$.

Mean rule validity analysis. For subjects in the two observation groups the combined mean rule validity was .581. A regression analysis yielded a line with a slope of 1.06 and intercept of -0.01, $r = .72$. The computed slope was no different than the unit slope, $t(1) = -0.36$, $p > .7$; the intercept was not significantly different from zero, $t(1) = -0.14$, $p > .8$.

The mean rule validity for the rule-training group was .76. The regression analysis produced a line with a slope of 1.38 and intercept of -0.19, $\underline{r} = .72$. The slope was significantly different from the unit slope, $\underline{t}(1) = 3.15$, $\underline{p} < .01$; the intercept was marginally different from zero, $\underline{t}(1) = -2.06$, $\underline{p} < .06$.

Questionnaire analysis. Fifteen of the 22 subjects in the rule-training group correctly identified all 5 rules. Four subjects included all 12 letters related to the rules but incorrectly included others. On average, rule-training subjects correctly identified 11.7 of the letters and incorrectly included only 0.3 letters.

Discussion

As expected, there were no differences in percent correct classification judgments between the two observation groups. The critical findings in this experiment were the lack of effects between observation groups in both the recognition reporting task and the recall test. In constrast with the predicted outcome, the rule-discovery group was no more successful at inducing and reporting important rules in the post-study phase questionnaire than the memory subjects. Recall that the rule-discovery subjects were specifically informed that the rules used to order the letter strings dealt with positions of letters within strings and the appearance of certain letters next to other letters. Perhaps their failure to correctly induce many of the possible rules was due to the large number of

hypotheses that could have been generated and tested. Follow-up studies to detect dissociations between verbal reports and classifications would include taking verbal protocols during the study phase for these subjects to discover what kinds of hypotheses they entertain (e.g., Mathews et al., 1989).

There was also no difference between observation groups in computed mean rule validities and the slopes generated by the regression analyses. While the mean rule validities in each case slightly underpredicted the percent correct classification judgments, there was no difference between the groups and no effect on the slope of the regression line. This failure to produce any progressive underprediction by mean rule validities of percent correct provides more evidence of the "robustness" of the effect discussed in Experiments 1a and 1b: more research needs to be done before the task can be used to dissociate recognition and classification.

The so-called "instructional set" effect has been difficult to replicate (Dienes, Broadbent, & Berry, 1991; Mathews et al., 1989; Millward, 1980; Reber et al., 1980, Experiment 2). The lack of a consistent, predictable outcome speaks directly to the issue of using instructions to evoke a particular learning set. It is difficult to know what subjects are doing as they observe the stimuli. Subjects under rule-discovery instructions may be working under erroneous, biased, inefficient strategies, or abandon

the search for rules entirely (Lewicki, Hill, & Sasaki, 1989). Those under memory-based instructions may realize, however imperfectly, the rule-ordered nature of the letter strings. This may explain the many "no effect" studies in the implicit learning literature. In light of this and other failures to replicate Reber (1976), discussed earlier, it should be concluded that the "instructional set" effect is not very useful for exploring dissociations between classifications and verbal awareness, at least in unspeeded test trials (cf. Turner & Fischler, 1993).

The difference in percent correct judgments between the observation groups and the rule-training subjects was expected, and essentially replicates Turner and Fischler (1993). Of interest in this study are the differences in the two reporting tasks.

It is no surprise that the subjects in the rule-training group, who studied the logical rules prior to taking the questionnaire, were able to complete the questionnaire so easily. In contrast with the observation groups, the regression analysis of the rule-training group revealed a substantial difference between the regression line and the unit slope. However, little can be made of this outcome. The rule reporting technique permitted subjects, many of whom scored 100% correct on the judgment task, to report featural rules (such as the first letter of a line) that had only moderate predictive validity. This simple characteristic of the reporting technique reduced the

close association between percentage correct and mean rule validities. As an instrument for predicting the classification performance of subjects who learn the set of logical rules, the regression analysis turned out not to be appropriate.

It is claimed that "implicitly" instructed subjects in the artificial grammar tasks make judgments in an "intuitive" fashion, whereas "explicitly" instructed subjects make judgments consciously and deliberately. Yet, for all the research done on implicit learning in these tasks, there has been no demonstration that explicitly instructed subjects can verbally articulate anything more than implicitly instructed subjects. A compelling demonstration of independence between classification and the ability to verbally justify classification judgments in artificial grammar tasks has yet to appear (Dulany, 1991). If the issue of interest is one of verbal awareness, then it is contingent upon the implicit knowledge "side" to develop adequate measures of verbal ability.

The failure in this experiment to dissociate report and classification performance based on the instructions variable motivates the search for a design in which classification performance cannot be explained so easily by the recognition of member features. It has been suggested that the artificial grammar task may simply be unsuitable for demonstrating dissociations between verbal performance and classification judgments (Broadbent, FitzGerald, &

Broadbent, 1986). "If one is trying to show the less common kind of discrepancy, improving verbal knowledge without a change in the quality of decision, then from the armchair the more academic tasks of concept and language learning do not seem very suitable. One can hardly imaging that a person who can define a concept would nevertheless be unable to pick out instances of it . . ." (p. 35). The challenge is to find a set of conditions that do demonstrate the phenomena of interest, and to introduce these methods into the artificial grammar tasks. The next experiment includes a classification test of "transfer" that is not easily explained by recourse to simple recognition of item features.

Note

[1]Although little was made of the fact by the authors, the equation is a nonlinear, dynamical equation, in which the output of one trial, or iteration, depends on the output of the previous trial. Interest in humans' ability to learn nonlinear sequences is a fairly recent phenomenon (cf. Neuringer & Voss, 1993).

CHAPTER 5

EXPERIMENT 3

Categories and category memberships are based on similarity between and within category members. A single word, "similarity," disguises the very difficult question of what it means for an object to be "like" another object, in the sense that the two may be treated as members of a single category; many theories of similarity assume that members can be treated as points in coordinate space (Tversky, 1977). Many theories of categorization begin with some discussion of features and the relations between features of category members, and the rules for weighing the relative contributions of each. A large number of detailed, computational accounts have been proposed to explain human categorization behavior; all can account for behavior in some particular domain, but all suffer in comparison to their competitors when shifted from their domain of interest (see Estes, 1986, for a review). There is still much work to be done on the acquisition of knowledge in simple domains as used here, for despite the effort that has gone into this work, scientists have reached no consensus with regard to categories and concepts (Medin, 1989).

The issue over the representation of knowledge in the artificial grammar task shares with other concept formation

tasks a substantial lack of agreement. Dulany et al. (1984, 1985) and Reber et al. (1985) described subjects' knowledge of the relations between features as one of "correlated grammars": Incomplete, often inaccurate knowledge of bigrams and trigrams within letter strings. Still unresolved is whether this knowledge is conscious and explicit or whether it is abstract, and unavailable to conscious awareness. The precise form of the knowledge acquired in the artificial grammar tasks has become one of renewed interest, due primarily to a series of experiments by Perruchet and Pacteau (1990) and Mathews et al. (1989).

Perruchet and Pacteau (1990), in a series of artificial grammar experiments, set out to demonstrate that performance in these tasks can be explained as nothing more than explicitly remembered letter pairs, or bigrams. They extended the Dulany et al. (1984) study by modifying the learning phase of the artificial grammar task. They also addressed the issue of the underlying form of the knowledge acquired in the task.

In their Experiment 1, subjects who studied a set of permissible bigrams performed as well in a classification test phase as subjects who studied complete letter strings. In Experiment 2, subjects made judgments of letter strings whose ill-formed strings contained violations consisting of invalid bigrams or of valid bigrams in nonpermissible locations. Judgments of strings composed of invalid bigrams was accurate; valid bigrams in nonpermissible locations was

extremely poor. The results indicated that memory for pairwise features was the critical factor in the task.

In Experiment 3, subjects took a recognition test of bigrams presented during a study phase; there was no difference between rule-discovery and memory groups. The bigrams and their collected recognition scores formed the basis of a simple simulated model of grammaticality judgments. Test items presented in Experiments 1 and 2 were judged ungrammatical if they contained an unfamiliar (low recognition score) bigram; otherwise they were judged grammatical. The judgment data from this simulation seemed to follow the data collected from human subjects in Experiments 1 and 2. Perruchet and Pacteau concluded that there was no need to posit knowledge of the structure underlying the formation of valid letter strings; explicit, fragmentary knowledge of valid letter strings accounted fully for classification performance.

The experiments by Perruchet and Pacteau suffered from some methodological problems that limit their usefulness (Reber, 1990). Some of the data collected in the first two experiments were eliminated based on some questionable posthoc analyses: They deleted test items that were nongrammatical based on impermissible first letters in order to make the results conform to the outcome of their simulation in Experiment 3. Also, the decision rule used in the simulation to decide if a bigram was unfamiliar appeared arbitrary and of questionable validity (Reber, 1990).

In these experiments, Perruchet and Pacteau were able to show that abstract knowledge of the set of items was not necessary for above-chance classification performance in a typical artificial grammar task. However, the classification task does not measure all of the information a subject might have acquired during the study of grammatical letter strings. Humans are opportunistic problem solvers in the sense that they take advantage of the information at hand to solve a given problem. For the tasks designed by Perruchet and Pacteau, it is entirely reasonable to view classification judgments as the result of remembered bigrams. Turner (1992) simulated classification judgments of the artificial grammar task using as a knowledge base only a small set of letters in valid positions and letter bigrams; the simulation performed as well as the best human subjects (80% correct). This would seem to support the simulation carried out by Perruchet and Pacteau.

However, to say that subjects can perform a task or solve a problem in a particular way does not preclude the existence of other approaches. Given the proper testing procedure, the study of grammatical, whole-item letter strings should produce an advantage in classification (or decision time, or some other dependent measure) when compared to study of rule-ordered letter pairs. Such an outcome would demonstrate that there is more learned in the artificial grammar task than simple memory of letter pairs.

In the artificial grammar tasks dependent measures such as anagram solving (Reber & Allen, 1978), concurrent verbal protocols (Mathews et al., 1989), and reaction times (Turner & Fischler, 1993) have all been used with success in an effort to uncover and describe the knowledge acquired by subjects in these tasks. Perruchet and Pacteau's designs tested for the explicit recognition of item features, not for the knowledge of underlying structure across a set of related items, and they found what they were looking for (Mathews, 1990). As shown in Experiment 1b, the recognition of a feature relies on its frequency of occurence in a set of studied items. The structure of a set of items, as implied by the grammar used to order the items, describes the relations between symbols within each item and within the set of items. A subtle, but more direct, test of the acquisition and knowledge of structure is the ability to transfer knowledge from one task to another that is structurally similar, but superficially different. Such studies of "transfer" in artificial grammar tasks have been used to test subjects' knowledge of the structure of the stimulus domain, rather than for their explicit recognition of features.

Reber (1969) tested subjects' ability to memorize and reproduce letter strings ordered by a finite-state grammar. In a subsequent transfer phase, subjects in the "symbols" group reproduced letter strings with the same grammar but different letters. Subjects in the "syntax" group

reproduced letter strings with the same letters ordered by a different grammar. The "syntax" group showed more impairment in the memory task than the "symbols" group. Reber claimed that subjects were learning the structure of the stimuli rather than groups of explicit symbols.

A similar effect was demonstrated in a classification task. Mathews et al. (1989) showed that subjects can transfer knowledge of an artificial grammar from one set of letters to another set of letters in a classification task, though imperfectly. Subjects were assigned to one of four groups based on instructions (memory or rule-discovery) and letter-set (same or different). In what the researchers called a "teach-aloud" procedure, subjects in the rule-discovery groups provided concurrent verbal descriptions of the rules by which they selected strings in a multiple choice design. Feedback was given as to the correctness of the judgments. During breaks between blocks of trials in the test phase, the memory groups also were asked to recall grammatical strings, or portions of strings from the study and test phases.

The experiment was conducted in 4 sessions over a period of 4 weeks. The same-letter-set groups studied and tested on the same set of letters throughout the first three sessions; in Week 4 (the transfer condition) subjects made discrimination judgments of items composed of a new, unfamiliar letter set. The different-letter-set groups studied and tested on different sets of letters over the

first three sessions, then transferred to a new test letter set in Week 4. The main finding was that performance as measured by selection of grammatical strings was significantly higher for all subjects in Week 4 than a group of no-study control subjects, who scored no better than chance. The Week 4 transfer manipulation demonstrated that subjects were not simply recognizing previously studied bigrams; they were making classification judgments of items constructed of letter sets they had never seen.

These results showed more than explicit recognition of letter pairs; they demonstrated the ability of subjects to apply knowledge of structure to new stimuli and new situations. The data also suggested that subjects integrated knowledge of structure across strings. Commenting on the design of the study phases in Perruchet and Pacteau (1990), "If their subjects' knowledge consisted almost entirely of pairwise associations, it would be utterly useless in a transfer task using the same grammar instantiated with an entirely different letter set" (Mathews, 1990, p. 415).

If subjects can acquire and apply structure from one domain to another in transfer tasks, under which conditions will transfer be facilitated? The Mathews et al. (1989) study raised some interesting questions. Significantly, there were no differences between groups in the final week of testing; neither the instructions nor the letter set manipulations seemed to affect transfer. Perruchet,

Gallego, and Pacteau (1992) pointed to the lack of effect based on instructions in criticizing the experiment as evidence for abstractive processing. If the "strong view" of implicit learning is true, shouldn't memory instructions produce better performance on the transfer task than rule-discovery instructions? This is a reasonable critique of the Mathews et al. study. However, the failure of one variable, such as instructions to subjects, does not effectively rebut the notion of abstraction in these artificial grammar tasks. The weakness of the instructions manipulation in producing differential effects in these tasks has already been discussed. Suffice it to say, the trick in these experiments is to find a set of conditions which demonstrate an effect, rather than a set of conditions in which no effect is produced.

The lack of an effect based on the letter set variable during training is more interesting. This result was interpreted by Mathews et al. as evidence for "automatic abstraction;" the acquisition of structure occurs at the same rate no matter what the surface features of the stimuli. Thus, the different letter sets were no more effective in training subjects for the transfer task than repeated exposure to the same letter set. This interpretation ignored the fact that subjects in the task had to study far longer than subjects in previous artificial grammar tasks in order to achieve above-chance levels of performance in classification tasks: about 30 minutes over 4

sessions compared with about 7 minutes in the typical artificial grammar experiment. If abstraction were indeed automatic then Mathews subjects should have reached asymptotic performance in 7 minutes (or less).

This interpretation also does not touch on an interesting feature of the reported data. In fact, the same-letter-set groups improved substantially in the discrimination task from Weeks 1 to 3, but fell substantially during the transfer manipulation (Week 4). The different-letter-set groups, in contrast, performed more poorly than the same-letter-set groups: They improved only slightly from Weeks 1 to 3, but continued to improve on the transfer task, performing up to the level of the same-letter-set groups. This apparent interaction between letter set and Week number was not tested, however, and Mathews et al. maintained that there were no differences between the letter-set groups.

The issue of letter sets and the abstraction of knowledge in artificial grammar tasks follows a similar issue in earlier studies in pattern recognition, namely, the importance of the variability of study items in classifying novel items. Posner and Keele (1968) tested two groups of subjects classifying novel dot patterns; subjects who studied items that were closely similar to a computed prototypical pattern were less accurate in classifying novel patterns than subjects who studied more variable items. Dukes and Bevan (1967) asked subjects to study faces;

subjects who studied faces of high-similarity were more accurate in recognizing previously studied faces than a low-similarity group. The low-similarity group, however, was more accurate in classifying new faces. The results of these studies and others on pattern recognition (c.f. Fried & Holyoak, 1984; Reed, 1972) argue against the automatic abstraction of structure independent of the features comprising the study items. In light of past research, variability in the sets of study items should enhance classification judgments in artificial grammar transfer tasks (Brooks & Vokey, 1991).[1]

There were two objectives to this experiment. The first objective was to explore the differences in classification performance between subjects trained on letters paired for presentation with subjects trained on the more typical display of grammatical letter strings. The first part of the experiment replicated Perruchet and Pacteau's (1990) Experiment 1, which compared subjects in a pairs-display group with a strings-display group. Additionally, the present study also extends the work of Perruchet and Pacteau by including a transfer test phase, in which subjects made classification judgments of letter strings generated by the same grammar but with different letter sets. It was hypothesized that the strings-display condition would produce better performance than the pairs-display condition in the initial, no-transfer classification trials. It was hypothesized further that subjects in the

strings-display condition, having abstracted knowledge of the structure of the strings, would be able to accurately classify novel items in the transfer task. Subjects in the pairs-display condition, having developed no knowledge of the structure of the grammar across strings, would be unable to perform the transfer task at a rate better than chance.

The second objective of this experiment was to study the effects of different letter sets in both the no-transfer and transfer test trials. Although Mathews et al. (1989) found no effect of changing letter sets on subjects' performance on a 5AFC version of the transfer task, earlier work on pattern recognition suggested that increasing the variability of the stimuli during study should improve performance during transfer. It was hypothesized that subjects in the different-letter-set condition would be more accurate in the transfer test than subjects in the same-letter-set condition. The 2 x 2 factorial design of the experiment also allowed analysis of the combined effects of letter sets and display presentation.

Method

Subjects and Design

The subjects were undergraduates at the University of Florida who participated as part of a requirement for an introductory psychology class. Eighty-two served in the experimental conditions, 15 in a no-study control group. The data from nine other subjects were discarded for making the same response on every trial in one or more trial

blocks. Subjects in the experimental conditions were assigned to one of four groups based on Display Type (strings or pairs) and Letter Set (same or different). The experiment was conducted in three trial blocks, which was a within-subjects variable.

Materials

This experiment employed the same finite-state grammar used in Experiment 2 (refer to Figure 5). The same items were used during the test phase; in addition, three new sets of test items were created by substituting a unique letter for each letter in the original letter set comprised of the letters PTKXS (henceforth referred to as Letter set 1). The following letters comprised the new letter sets: Letter set 2, BLYFC; Letter set 3, NHMZJ; Letter set 4, DGQRW. Thus, the rules used to order the letter strings remained the same, but there were four unique letter sets employed as stimuli. The stimuli were presented and all responses taken on IBM-compatible microcomputers.

Strings Display Type. The Display Type variable refers to the presentation of study items during the study phases of the experiment. The strings-display used the same 21 study items that were used in Experiment 2. As with the test items, three new sets of study items were created by substituting three new letter sets for the letters in Letter set 1. The entire set of strings-display study items and the test strings are seen in Appendix E. None of the study items appeared in the set of test items.

Pairs Display Type. The pairs-display was created by showing the letters used in the study phase in pairs, rather than as entire strings as in the previous experiments. The set of letter strings in the study set comprised 145 letters, or 72.5 pairs of letters. To facilitate display, 70 pairs of letters were chosen for the study set. The pairs were chosen so that the frequency of occurrence of each pair best matched its frequency of occurrence in the letter strings from which it was extracted. For example, pair KP appears 14 times in the study set in the total of 124 pairs (124 is the reference here because each letter of a string except the first and the last ones, enters into two different pairs). Therefore, KP was presented (14/124) x 70 = 7.9 times, rounded off to 8 times, to the pairs groups. This procedure follows closely the procedure from Experiment 1 in Perruchet and Pacteau (1990). The pairs of letters from Letter set 1, and their frequency of occurrence in the study set, are seen in Appendix F. Like the strings-display, the study pairs were constructed with four letter sets.

Procedure

In the study phase of each trial block, subjects in the strings groups viewed three grammatical letter strings in each of seven screens. Each screen was displayed for 30 seconds after which the screen went blank. Letter strings of similar appearance were distributed across frames to avoid introducing structure that would make the grammar

"salient." Subjects in the pairs groups viewed 10 pairs of letters in two columns in one of seven screens. Each screen was displayed for 30 seconds. Total viewing time of the study items was 3.5 minutes for both the strings and pairs groups. All subjects were asked to study the items carefully, and told that they would be asked some questions about the items.

In the test phase of Trial Block 1, subjects in all four experimental groups learned that the items were ordered by a set of rules. All subjects made classification judgments of 40 novel, randomly ordered letter strings presented one at a time on the computer screen. Subjects made judgments by pressing "1" on the keyboard to indicate a grammatical judgment and "0" for a nongrammatical judgment. Between each trial block a message indicated that subjects could take a break before continuing the experiment.

In Trial Block 1, all test phase items used the same letter set as the items in the study phase. In Trial Block 2, subjects in the different-letter-set groups studied and tested on a different letter set from the letter set in Trial Block 1; subjects in the same-letter-set groups studied and tested on the same items from Trial Block 1. Trial Block 3, the "transfer" block, was based on the transfer manipulation in Mathews et al. (1989). The same-letter-set groups studied items of the same letter set as in Trial Blocks 1 and 2, but tested on a new letter set (their second). The different-letter-set groups saw a new letter

set during study and another new letter set during the test (their fourth). The presentation of letter sets was counterbalanced across subjects.

Control Condition

Subjects in the control condition took part only in the test phases of the experiment. They were informed that some of the test items were based on a set of rules, and to make their judgments based on whether they thought the items appeared to follow rules. To justify this difficult task, it was emphasized to these subjects that the experiment was a test of "intuition" rather than problem solving ability. Each subject saw a different letter set in each trial block. The presentation of letter sets was counterbalanced across subjects.

Results

The mean judgment accuracy and standard deviations for each cell in the 3 x 2 x 2 design (Trial Block x Letter Set x Display Type) are presented in Table 3. The mean score for Trial Blocks 1 and 2 are shown under the heading "Combined 1&2." The first analysis compared the percent correct judgments for the four experimental groups across all three trial blocks.

The different-letter-set groups were significantly more accurate in classifying the test items than the same-letter-set groups, 58% and 55% respectively, $F(1, 78) = 9.94$, $p < .005$. The strings-display group scored higher than the pairs-display group, 57.8% and 55.2%, respectively, $F(1, 78)$

= 7.48, $p < .01$. The Letter Set x Display Type interaction was not significant, $F(1, 78) < 1.0$.

Table 3

Mean Percent Correct Judgments by Letter Set and Display Type for the Experimental Groups in Experiment 3.

		Trial Block			
Letter Set	*n*	1	2	Combined 1&2	3
		Strings			
Same	22				
M		56.0	58.5	57.3	54.3
SD		6.9	7.6	4.4	7.4
Different	20				
M		59.8	62.4	61.1	56.0
SD		6.9	8.3	6.0	6.9
		Pairs			
Same	19				
M		51.7	58.0	54.9	50.7
SD		5.9	8.1	4.0	6.6
Different	21				
M		56.1	60.5	58.3	53.6
SD		7.8	10.3	6.3	6.9

The analysis also revealed a main effect of Trial Block, $F(2, 256) = 14.7$, $p < .0001$. Trial Block did not interact with Letter Set or with Display Type, all Fs < 1.0.

Analysis of No-Transfer Trial Blocks

The next analysis was concerned with the effects of the variables across Trial Blocks 1 and 2 only. The percent correct judgments were analyzed by a 2 x 2 x 2 ANOVA.

The different-letter-set groups were significantly more accurate in classifying the test items than the same-letter-set groups, 59.6% and 56.2% respectively, $F(1, 78) = 9.43$, $p < .005$. The strings-display groups were significantly more accurate than the pairs-display groups, 59.1% and 56.7% respectively, $F(1, 78) = 4.91$, $p < .05$. The Letter Set x Display Type interaction was not significant, $F < 1.0$.

The general trend was for subjects to improve on the task between the no-transfer trial blocks; there was a main effect of Trial Block, as subjects scored higher on Trial Block 2 than on Trial Block 1, 59.8% and 55.9% respectively, $F(1, 78) = 9.61$, $p < .005$. Trial Block did not interact with Letter Set or with Display Type, all Fs < 1.3.

Analysis of Transfer Manipulation

The next analysis of the experimental groups compared the percent correct judgments for each of the four groups on Trial Block 3, the transfer trial block. The strings-display group scored marginally higher than the pairs-display group, 55.1% and 52.2% respectively, $F(1, 78) = 3.88$, $p < .06$. There was no effect of Letter Set and no

Letter Set x Display Type interaction. Duncan's Multiple Range test shows the relative ordering of the four groups; the clusters are shown next to the means in Table 4.

Table 4

Duncan's Multiple Range Test on percent correct judgments in Trial Block 3.

Experimental groups	Mean	Grouping	
Different strings	56.0	A	
Same strings	54.3	A	B
Different pairs	53.6	A	B
Same pairs	50.7		B

Note: Groups that share a common letter are not significantly different.

Table 5

T-tests run for Groups in Trial Block 3, Better Than Chance Performance.

Experimental groups	df	*t*	*p*
Different strings	19	3.91	0.001
Same strings	21	2.72	0.013
Different pairs	20	2.37	0.028
Same pairs	18	0.43	0.67

Finally, individual t-tests were conducted to test for above-chance performance for the four groups. The results are shown in Table 5. Of the four groups, only the same-pairs group failed to achieve above-chance performance on the transfer test.

In response to the failure of the pairs groups on the transfer task, it might reasonably be argued that the no-transfer phases might have been more difficult for the pairs groups than the strings groups, resulting in a lower level of learning in the pairs groups prior to the beginning of the transfer phase. This is plausible, in light of the lower levels of performance by the pairs groups across Trial Blocks 1 and 2. In anticipation of this criticism, a final analysis concentrated on those subjects who demonstrated their having learned the "grammar" by individually scoring better than chance during Trial Blocks 1 and 2. Eight subjects in the pairs groups (seven in the different-letter-set condition and one in the same-letter-set condition) scored higher than chance (at least 49 out of 80, or 61.25% correct). Fifteen subjects in the strings groups (11 in the different-letter-set condition and 4 in the same-letter-set condition) scored higher than chance.

The data from these 23 subjects were analyzed by a 2 x 2 x 2 (Trial Block x Letter Set x Display Type) ANOVA. The 15 selected subjects in the strings groups scored no higher than the 8 subjects in the pairs groups, 64.8% and 64.4%, respectively. There were no main effects of Letter Set,

$F(1, 19) = 1.2$, or Display Type, $F(1, 19) = .75$, and no interactions. There was an effect of Trial Block, $F(1, 19) = 8.75$, $p < .01$.

Mean percent correct on the transfer task for the 8 selected subjects in the pairs condition was 54.1% (SD = 7.1). A t-test revealed that the score was no greater than chance, $t(7) = 1.63$, $p > .1$. The 15 subjects in the strings condition scored 57.2% correct (SD = 7.6), which was significantly higher than chance, $t(14) = 3.68$, $p < .01$.

Control Condition Analysis

The mean judgment accuracy and standard deviations for the no-study control group are presented in Table 6. The data were analyzed by an ANOVA.

Table 6

Mean Percent Correct Judgments for the Control Group in Experiment 3.

	Trial Block			
	1	2	3	Total
M	46.3	51.7	51.3	49.8
SD	6.6	4.8	4.2	2.7

Note: $n = 15$.

There was a main effect of Trial Block on judgment accuracy, $F(2, 28) = 4.33$, $p < .05$. Planned t-tests showed that the control subjects scored higher on Trial Block 2 than on Trial Block 1, $t(15) = 2.4$, $p < .05$; there was no difference in accuracy between Trial Blocks 2 and 3, $t(15) = -0.18$, $p > .8$.

Subjects' scores ranged from 45% correct to 54.2% correct. Mean accuracy over all three trial blocks was only 49.8%. This confirmed that the relatively higher levels of performance by the experimental groups was due to learning during the study phases of the experiment rather than to any inherently detectable differences between the grammatical and nongrammatical letter strings.

Discussion

Within the no-transfer trial blocks, the hypothesized superiority of the strings training over the pairs training condition was confirmed. The outcome failed to support Perruchet and Pacteau's (1990) claim that the knowledge acquired in the task consists strictly of pairwise comparisons. It should be emphasized again that this design was not one of simple recognition of exemplars; none of the study items appeared in the test set in the strings condition. The inclusion of study items in the test set might have conferred an obvious advantage on the strings group, as subjects could have memorized and recognized the study items, making classification relatively easy. Instead, the advantage seemed to derive from the display of

entire strings, and the subjects' ability to integrate information about pairs of letters across entire strings.

It had been suggested that the pairs display would be "utterly useless" in a transfer test of classification judgments (Mathews, 1990). This hypothesis was also largely confirmed. The same-pairs group, the group most like Perruchet and Pacteau's (1990) pairs group, scored no higher than chance in the transfer test. Likewise, the pairs groups when analyzed together scored no higher than chance. This was in contrast to the strings groups, each of which scored higher than chance in the transfer test.

It could be argued that the no-transfer trial blocks were more difficult for the pairs groups than the strings groups, which may have resulted in there being nothing to transfer during the transfer test. That is, were the two no-transfer trial blocks less effective in training the subjects prior to the start of the transfer trial block, making the transfer block harder for the pairs groups than the strings groups? This is a reasonable observation, given that only seven subjects in the pairs groups individually scored higher than chance during the no-transfer trial blocks. Yet even these seven subjects, who performed as well or better than the average subject in the strings groups, were unable as a group to score higher than chance in the transfer test. This result, as well as any, showed the pairs training as "utterly useless" in a test of transfer in this task.

There was a small tendency for above-chance performance in the transfer test for the different-pairs group. But, as with the same-pairs group, there was a possible confound. All subjects saw grammatical and nongrammatical strings during the test phases of the no-transfer trial blocks, which may have contributed to their ability to perform the transfer test. Recall that the control subjects, who viewed and classified only test strings, increased from 46.3% to 51.7% between Trial Blocks 1 and 2, a small but statistically significant amount. Thus, the act of judging the test strings during the no-transfer trial blocks may have improved performance slightly during the transfer test phase.

Within the no-transfer trial blocks, the superiority of different-letter-sets versus the same-letter-sets was somewhat unexpected. Mathews et al. (1989) found that the same-letter-set groups gradually improved on the classification task during the no-transfer trial blocks, relative to the different-letter-set groups, but fell substantially during the transfer test. The different-letter-set groups, on the other hand, showed only a small improvement between no-transfer trial blocks, but continued to improve during transfer. The differences between the tasks could account for the different results; Mathews subjects chose from four visible alternative items on each trial, a design which emphasizes comparison of a few items within each trial. The yes-no decision made for each item

in the present experiment may emphasize memory for structure over the entire set of items. That is speculation, but it is clear that the task relies on a combination of knowledge of structure and explicit recognition of letter bigrams and trigrams. So, it might have been found that performance for the same-letter-set groups would increase at a faster rate during no-transfer trial blocks, as seen in Mathews et al (1989). This was not the case, as the different-letter-set groups outperformed the same letter set groups in both the no-transfer and transfer tests. Such a finding indicates the greater relative importance of structure compared to recognition of features, at least when subjects are faced only with decisions about individual items.

The results of the transfer test clearly indicate the importance of different training sets over the same letter sets. The hypothesis that training with different letter sets would result in better classification performance in the transfer task was confirmed. This finding, when taken together with the transfer data in Mathews et al. (1989), showed that the acquisition of structure is not fully "automatic," as suggested by Mathews et al. (1989), but is tied to conditions during study. More important than the main effect of letter set, however, was the results of the four conditions: the different-strings group scored the highest on the transfer task, demonstrating the combined advantage of both the different letter sets during study and the strings presentation. As pointed out earlier, the same-

pairs group scored the lowest of the four groups, and were unable to score higher than chance. The relative ordering of the four groups demonstrated the contributions of letter set and display type presentation on performance of the transfer test. Further testing to detect possible interactions between the display type and letter set variables would make the test items easier to discriminate; the scores on the transfer test ranged from just 50% to 56%.

Note

[1]The idea that variability is an important determinant in concept formation apparently goes back to the notion of "learning sets" in the animal learning literature (Harlow, 1949).

CHAPTER 6

GENERAL DISCUSSION

Experiment 1a replicated two well-known studies by Dulany et al. (1984, 1985) that purported to show that reported featural rules correlated with, and hence were directly responsible for, classification judgments in artificial grammar tasks. These studies have been cited frequently in support of a fully "explicit" model of categorization, and against the notion of implicit, abstract knowledge. Experiment 1b simulated the generation of rule reports in the task, and demonstrated the possibility that reports and classification judgments could be generated independently of one another to produce the close association between reports and judgments found in Experiment 1a. The results showed the inadequacy of the regression analysis on reports and judgments for distinguishing between the two models of performance, and for detecting dissociations between recognition and classification in the task.

The design of Experiment 2 followed that of Experiment 1a, but modified the report procedure in such a way as to force subjects to be more specific about their rule reports. This modified report procedure failed to alter the close correspondence between the computed rules validities and

percentage correct judgments and demonstrated the difficulty of dissociating recognition and classification measures in this task.

In Experiment 3, the classification test included a transfer test in which subjects made judgments of items constructed of a novel, previously unstudied letter set. The experiment clearly demonstrated that classification judgments could be made in the absence of explicit recognition memory of studied items; the test items in this case shared no surface features with the studied items. The results were offered in support of the hypothesis that subjects were abstracting the underlying structure of the stimuli both within and across studied items and using knowledge of that structure to correctly classify the novel test items. The experiment also demonstrated the importance of variable stimuli and whole items during study for developing an abstract representation of the stimulus domain.

What is Abstraction?

Thus far, the term abstraction has been used without giving a precise definition. In what sense might knowledge be considered abstract? One line of evidence for abstract knowledge comes from tasks that show subjects will readily accept as familiar novel patterns or items that fall close to a computed prototype or central tendency of a group of items. This "family resemblance" notion of abstraction is common in the literature of categorization and concept

formation (Smith & Medin, 1981). Another line of evidence for abstract knowledge comes from work in problem solving in which subjects are able to see beyond dissimilarities in the details of problem domains and solve problems in superficially different but structurally similar domains. These two apparently dissimilar meanings for the same term demonstrate the range of the phenomenon, but also give an indication of the imprecision of its definition. The approach, popular among connectionists, that problem solving is just a special form of pattern recognition (see Holyoak & Thagard, 1989) does little to help clarify issue of abstraction.

Perruchet and Pacteau (1990) point out, in criticizing theories that rely on abstraction, that our concept of abstraction is loosely defined. Nevertheless, it does not mean that the idea is without merit or has not been used with success in describing knowledge which goes beyond explicit recognition. There is little to be gained by simply describing an idea as "loosely defined," and dismissing any discussion of it. Prior to the pioneering work of Treisman and Broadbent in the 1950's and early 1960's, ideas about attention were also loosely defined. Likewise, intelligence, awareness, and consciousness are not fully defined, yet these concepts are freely used in the literature. The point of research on transfer is to shed light on the (presently) loosely defined concept of abstraction.

Mathews (1990), in an effort to better define the characteristics of abstraction, suggested two operational definitions of abstraction in these artificial grammar tasks. Abstraction could refer to (a) the level of generality of the rules that correctly describe letter strings, or exemplars and (b) the ability to make correct classification judgments of novel letter sets. The first definition (a) concerns the generality of a featural rule that could include an exemplar. To the extent that the featural rule was inclusive of a greater numbers of exemplars the rule could be said to be more abstract. Thus, the rule "select strings that begin with VXT" is more abstract than the rule "select strings that begin with V." This definition of abstraction recalls the use of the term by Rosch (1978), in reference to category levels. Superordinate categories, like "furniture," are considered more abstract than basic or subordinate categories in the sense that there are few invariant features that include all members of the category. It is easy to see a superficial resemblance between this idea about abstraction and the rule "strings that begin with V" as a sort of an abstract, superordinate category.

This tacit equivalence between the abstraction found in the superordinate categories of natural phenomena and the featural rules of artificial stimuli seems rather tenuous. Superordinate categories, and membership within those categories, often depend on context rather than perceptual

characteristics. For example, many items commonly considered basic members of the furniture category can function as members of the superordinate category "weapons" under the right conditions. Context, of course, is absent in the artificial grammar tasks; the abstraction implied by general rules, such as "strings that begin with a V" are strictly perceptual.

The second definition (b) of abstraction suggested by Mathews refers to the transfer task. This sense of transfer is more generally associated with research in problem solving, in which subjects must solve difficult reasoning problems after having been exposed to superficially different but structurally similar problems and their solutions (Gick & Holyoak, 1980, 1983). In the artificial grammar task, as with the problem solving studies, "we say abstract knowledge was acquired when transfer to a relatively novel task is successful" (Mathews, 1990, p. 415).

The definition of abstraction as transfer to novel stimuli seems by contrast the more compelling example of abstract knowledge. What is being learned in the artificial grammar task are relationships between sets of symbols across a set of stimuli, rather than specific letter pairs. It is this knowledge that is acquired and displayed during the transfer task. These results are also the ones that exemplar based accounts of categorization have the most trouble with. Obviously, there are no exemplars to

recognize; all classification is of items that do not resemble study items.

Abstract Structure or Abstract Analogy?

Exemplar theories of classification have difficulty dealing with the results of transfer studies of artificial grammars. Obviously, categorization in the transfer tasks cannot be computed from explicitly recognized letter bigrams and trigrams. One effort to address this phenomenon of transfer in artificial grammars tasks from an exemplar viewpoint describes the process as one of abstract analogy, or relational analogy (Brooks & Vokey, 1991). As a means of classifying novel items, analogy relies on within-item relations of the features, rather than on the abstraction of the overall structure of the study items. For example, a subject might observe an items such as MXVVVM and BDCCCB and notice a common pattern: the same symbol at the beginning and end, and a symbol repeated 3 times in between. In this way, two general rules are abstracted from the first item, and applied to the second item in order to classify it. This approach to explaining transfer as one of analogical reasoning recalls an approach common to case-based theories of problem solving (Gentner, 1983; Gick & Holyoak, 1980; Kolodner, 1988)

In a study designed to measure the relative importance of knowledge of structure (the grammar) and specific similarity to study items, subjects studied grammatical letter strings, then took a transfer test of "near" and

"far" changed letter set items (Brooks & Vokey, 1991). Near items defined as those with just one letter difference from a studied item; far items were those that had more than one letter different from studied items. Subjects were better at classifying the near items than the far items. These data were presented in support of the importance of analogical reasoning; the process of drawing analogies between study and test items was easier with the items that closely resembled the study items than with the "far" test items.

Some questions about this account of transfer suggest themselves. First, if subjects are abstracting rules such as those described above, then verbal protocols taken during study or test should reveal those rules. None of the studies cited in support of the analogy perspective attempt to assess subjects' verbal knowledge of abstract rules. Perhaps the abstract rules are entirely unconscious and cannot be tested verbally; if this is the position the authors take they do not say.

The second question relating to the analogical position deals with the effect of the study of grammatical items in preparation for the test phase of the experiment. Brooks and Vokey clearly believe that much the knowledge acquired during study of grammatical items is not abstracted across the entire set of items, but consists mainly of relations between features within individual items.

In a direct test of this hypothesis, Wittlesea and Dorken (1993) trained subjects on items with salient patterns of repetition within each item (e.g., HDFX-HFDX), but no obvious between-item pattern. "Legal" test items were created from the study items using a novel letter set (e.g., GZTP-GTZP). "Illegal" test items were created by reordering legal test items (e.g., GZTP-GPTZ). Subjects were able to discriminate legal from illegal items, showing that abstraction of structure and transfer to novel stimuli can occur within items as well as between items. A second experiment looked at the effect of training on the ability to discriminate novel legal from illegal items. Subjects viewed 25 letter strings generated by a relatively unstructured artificial grammar; the items were made distinctive by letters that appeared two or three times within the item. Subjects who were directed to process the study items for the presence of letter pairs scored as well on a test of novel letter set items as on original letter set items. Two groups of subjects who processed the strings in an incidental manner were impaired on the transfer test. These data give support for the idea that relations within items can be abstracted and applied to novel stimuli.

This account of transfer as one strictly of analogy based on within-item structure would seem to reduce the relative importance of studying grammatical items prior to test. In fact, it would suggest that the "study" phase could consist of a typical test phase, in which subjects

make classification judgments, then transferred to a second test phase using items constructed of a new letter set. This was the design for the control subjects in the present Experiment 3, who scored less than 51% correct. The data from these control subjects indicate the importance of knowledge acquired by experimental subjects across a set of grammatical items during the study phase.

Although they downplay the relative importance of an abstracted grammar in these artificial grammar tasks, Vokey and Brooks (1992) found enough evidence for abstraction of structure to suggest that two processes may be at work in the transfer task. The model incorporates both abstraction of structure and abstract analogy as necessary for performance. In this way, they attempt to reconcile an exemplar position of performance in artificial grammar tasks with the strong abstractive views proposed by Mathews (1990) and Reber (1989). Their model also modified an earlier strong exemplar theory of performance (Brooks, 1978).

Two Approaches to Process Dissociation

In light of failure of Experiment 2 to dissociate classification from reporting measures, and some of the earlier difficulties with invoking a particular learning set, what is the best approach to take in uncovering separate processes? There are two lines of research worth mentioning in this regard: decision times analysis and work with amnesics.

<u>Speeded Judgments and Decision Times</u>

Decision times have been used with success to draw inferences about the underlying representation and dynamics of memory in lexical decision tasks (McClelland, 1979; Meyer, Irwin, Osman, & Kounios, 1988; Meyer & Schvaneveldt, 1971; Wickelgren, 1977). Until recently, however, the use of decision times in artificial grammar tasks to clarify the hypothesized relationship between implicit and explicit processes was unremarkable (e.g., Reber et al., 1980).

Turner and Fischler (1993) introduced a judgment deadline procedure to the standard artificial grammar design in an effort to dissociate implicit and explicit processes. The deadline procedure, adapted from Pachella (1974), required subjects to make classification judgments under long (6 seconds) and short (2 seconds) time pressure. In Experiment 1, subjects under study phase memory instructions showed no decrease in performance during the test phase between long and short deadline conditions; subjects under rule-discovery instructions dropped significantly due to time pressure. In Experiment 2, all subjects made "similarity" judgments of the test strings instead of the typical "rule-based" instructions; the differences between groups disappeared. The results demonstrated the usefulness of the deadline procedure for differences in processing based on instructions at time of study, an effect that has been difficult to produce.

In their Experiment 3, a group of subjects were explicitly trained on the set of logical rules needed to perfectly discriminate grammatical from nongrammatical strings (Turner & Fischler, 1993). These rule-training subjects performed near ceiling under long deadline conditions, but were significantly more impaired on short deadline than the groups that learned by observation. A detailed analysis of decision times showed that the rule-training group in judging the items applied the rules sequentially, while the observation learners responded in essentially a "wholistic" or parallel fashion. The results of this experiment were related to an hypothesis developed by Broadbent, FitzGerald, and Broadbent (1986). In their view, implicit and explicit processes may best be thought of as separate and distinct modes of processing, each with its own advantages. Explicit processing may be thought of in terms of a "look-ahead" strategy," which involves serial processes, the steps of which are verbally available. Implicit processing, which is faster and more appropriate for real-time situations, occurs by means of a "look-up table," in which situations and actions are stored and consulted as needed. The look-up table representation of knowledge of artificial grammars recalls some models of pattern recognition.

Classification and Recognition in Amnesia

Evidence for separate physiological systems in classification and exemplar memory was presented in a study

by Knowlton, Ramus and Squire (1992). Amnesic patients and subjects with unimpaired memories studied 23 grammatical letter strings one at a time for 3 seconds each. In a classification test, the normal subjects scored no higher than the amnesic patients. In the second test phase of the experiment, subjects were given a recognition test of study items; normal subjects were significantly more accurate in recognizing study items than amnesic patients. They argued that the results support the independence of declarative and procedural memory systems (see also Knowlton & Squire, in press).

Conclusion

The artificial grammar experiments are considered by some to be analogs of natural concept learning. The advantages of the artificial grammars are that they are tractable, understandable, can be easily manipulated, and prior knowledge is controlled for. They have several obvious disadvantages, though. They are not ordinarily as complex as natural concepts, and usually do not reflect the structure inherent in natural concepts. Furthermore, the amount of time devoted to their study is less; most studies using artificial stimuli take place in course of one day. This raises the issue of whether the results from these studies "scale up;" that is, can we take what is learned from the studies and apply that knowledge to natural concept learning? In terms of the present study it is recognized

that these artificial grammar tasks are letter-learning tasks standing in for the acquisition of natural concepts.

The research in artificial grammar learning does not touch on the more recent trend of discussing categories and their members in terms of their context (Medin, 1989; Medin & Shoben, 1988). In the present experiments, the influence of context is virtually nonexistent. The stimuli are related only by their surface features (the letters) and the relationships between features (represented by the finite-state grammar). Therefore, the results of these experiments do not address important issues related to the effects of context on categorization.

One underlying theme of this research is that a person can approach a task in more than one way. Indeed, a measure of intelligence is the ability to change one's representation of a given task. Interestingly, while we recognize that the ability to dynamically change representations given the constraints of some problem is a characteristic of an accomplished learner, there is less acknowledgment that scientific models should also behave dynamically (Lindsay, 1991). If scientists wish to accurately characterize human knowledge then their own mental models should reflect this ability. In order to adequately understand concepts and their acquisition, our representation of concepts must allow for multiple models of concepts and the human ability to change mental models of these concepts.

APPENDIX A
G1, G2, AND G3 CALCULATIONS OF MEAN RULE VALIDITY

The rule validity of a given featural rule is a proportion based on the number of times the featural rule appears in items of a selected category and the number of times the featural rule appears in all items in all categories. The featural rules are only counted against the items that are judged during the experiment (i.e., the test items). This is the equation for computing the rule validity of a featural rule in an item judged "grammatical" (Dulany et al., 1984, p. 547):

1. When the subject reports a rule asserting that "Feature *i* implies that String *j* is grammatical (G),"

$$\text{Validity} = P(S_j \in G \mid F_i \in S_j)$$

$$\text{Validity} = \frac{N(S_j,\ F_i \in S_j \text{ and } S_j \in G)}{N(S_j,\ F_i \in S_j)}$$

2. When the subject reports a rule asserting that "Feature *i* implies that String *j* is nongrammatical (NG),"

$$\text{Validity} = P(S_j \in NG \mid F_i \in S_j)$$

$$\text{Validity} = \frac{N(S_j,\ F_i \in S_j \text{ and } S_j \in NG)}{N(S_j,\ F_i \in S_j)}$$

For example, suppose that a subject has judged the item "MTV" as grammatical, and reports the item fragment "MT." MT appears in $\underline{x}$ grammatical items and $\underline{y}$ nongrammatical items. The computed rule validity for this item fragment is $\underline{x} / (\underline{x} + \underline{y})$.

There are different senses in which a fragment is judged to have appeared in an item as a feature, but the formula for computing the validity of a rule is general to any of the ways. Following Dulany, three different senses of "rule" are employed in the analysis: G1, G2, and G3.

G1: An item fragment, such as MT, is counted as being in an item if it appears anywhere in an item, regardless of its position in the item. If MT is reported as the feature in MTV, then the feature also appears in items such as MTTV and MMTV.

G2: The item fragment is counted as being in an item only if the fragment appears in the same initial position as

the reported featural rule. Following the example, "MT" appears in "MTV" and in "MTTV," but not in "MMTV."

G3: The item fragment is counted as being in an item only if the fragment appears in the same initial and final position as the reported featural rule. In other words, the rule validity is only computed against items of the same length. Thus, MT appears in MTV and MTM, but not in MTTV.

Dulany et al. (1984) found that the G3 computation, which is the most constrained calculation, yielded the best predictive validity. This is the most restrictive sense of feature, but seems to yield the best predictive validity. It is the calculation they used to report their data on p. 547. I reported mean rule validities using the G3 calculation in order to compare my results directly with those reported by Dulany et al.

On the other hand, the algorithm in Experiment 1b counted the presence of features from its first position in a string, regardless of the length of the string. This is similar to the way the G2 calculation counts features, which is more inclusive than the way the G3 calculation is made. This was done in order to make the simulation as unconstrained as possible and still account for the observed data.

APPENDIX B
LETTER STRINGS USED IN EXPERIMENT 1

Study	Test: Grammatical	Test: Nongrammatical
MTTTTV	VXTTTV	VXRRT
MTTVT	MTTTV	VXX
MTV	MTTVRX	VXRVM
MTVRX	MVRXVT	XVRXRR
MTVRXM	MTVRXV	XTTTTV
MVRX	MTVRXR	MTVV
MVRXRR	MVRXM	MMVRX
MVRXTV	VXVRXR	MVRTR
MVRXV	MTTTVT	MTRVRX
MVRXVT	VXRM	TTVT
VXM	MVT	MTTVTR
VXRR	MTVT	TVTXV
VXRRM	MTTV	RVT
VXRRRR	MVRXR	MXVT
VXTTVT	VXRRR	VRRRM
VXTVRX	VXTV	XRVXV
VXTVT	VXR	VVXRM
VXVRX	VXVT	VXRT
VXVRXV	MTV	MTRV
VXVT	VXRRRM	VXMRXV
	VXTTV	MTM
	VXV	TXRRM
	VXVRX	MXVRXM
	VXVRXV	MTVRTR
	MVRXRM	RRRXV

APPENDIX C
LETTER STRINGS USED IN EXPERIMENT 2

Study	Test: Grammatical	Test: Nongrammatical
TSSXS	TSXS	TTXS
TSSSSXS	TSSSXS	TKSSXS
TXXKK	TSSSSSXS	TSSSPPXS
TSSXXKK	TSXXKK	TSPPKK
TSSSXXKK	TXXTKK	TTXTKK
TXXTTKK	TXXTTTKK	TKKTTTKK
TSXXTKK	TSXXTTKK	SSXXTTKK
TXXKPXKK	TSSXXTKK	TSSXXTXX
TSXXKPS	TXXKPS	TXXKPT
TSSXXKPS	TXXTKPS	KXXTKPS
TXXTTKPS	TSXXTKPS	TSPPTKPS
PTTKK	PTKK	PXKK
PTTTTKK	PTTTKK	KTTTKK
PKPXKK	PTTTTTKK	PTTTTTKP
PTTKPXKK	PTKPXKK	PTKPPKK
PKPXTTKK	PKPXTKK	PSPXTKK
PKPS	PTKPXTKK	PSKPXTKK
PTTKPS	PTKPS	PTKPX
PTTTTKPS	PTTTKPS	PXTTKPS
PTKPXKPS	PKPXKPS	XKPXKPS
PKPXTKPS		

APPENDIX D
QUESTIONNAIRE: EXPERIMENT 2
OBSERVATION GROUPS

1. Which letter or letters appeared as the first letter in a line of letters (circle all that apply)?

P T K X S

2. Which letter or letters appeared as the last letter in a line of letters (circle all that apply)?

P T K X S

3. Which letter or letters appeared as doubles or triples? That is, which letters could be repeated in a line (circle all that apply)?

P T K X S

4. Which letters appeared as the first PAIRS of letters in a line of letters (circle all that apply)?

P P	P T	P K	P X	P S
T P	T T	T K	T X	T S
K P	K T	K K	K X	K S
X P	X T	X K	X X	X S
S P	S T	S K	S X	S S

APPENDIX D (CONTINUED)
QUESTIONNAIRE: EXPERIMENT 2
RULE-TRAINING GROUP

1. Which letter or letters could appear as the first letter in a line of letters (circle all that apply)?

P T K X S

2. Which letter or letters could appear as the last letter in a line of letters (circle all that apply)?

P T K X S

3. Which letter or letters could appear as doubles or triples? That is, which letters could be repeated in a line (circle all that apply)?

P T K X S

4. Which letters could appear as the first PAIRS of letters in a line of letters (circle all that apply)?

P P	P T	P K	P X	P S
T P	T T	T K	T X	T S
K P	K T	K K	K X	K S
X P	X T	X K	X X	X S
S P	S T	S K	S X	S S

APPENDIX E
LETTER STRINGS USED IN EXPERIMENT 3: SET 1

TSSXS	TSXS	TTXS
TSSSSXS	TSSSXS	TKSSXS
TXXKK	TSSSSSXS	TSSSPPXS
TSSXXKK	TSXXKK	TSPPKK
TSSSXXKK	TXXTKK	TTXTKK
TXXTTKK	TXXTTTKK	TKKTTTKK
TSXXTKK	TSXXTTKK	SSXXTTKK
TXXKPXKK	TSSXXTKK	TSSXXTXX
TSXXKPS	TXXKPS	TXXKPT
TSSXXKPS	TXXTKPS	KXXTKPS
TXXTTKPS	TSXXTKPS	TSPPTKPS
PTTKK	PTKK	PXKK
PTTTTKK	PTTTKK	KTTTKK
PKPXKK	PTTTTTKK	PTTTTTKP
PTTKPXKK	PTKPXKK	PTKPPKK
PKPXTTKK	PKPXTKK	PSPXTKK
PKPS	PTKPXTKK	PSKPXTKK
PTTKPS	PTKPS	PTKPX
PTTTTKPS	PTTTKPS	PXTTKPS
PTKPXKPS	PKPXKPS	XKPXKPS
PKPXTKPS		

LETTER STRINGS USED IN EXPERIMENT 3: SET 2

LCCFC	LCFC	LLFC
LCCCCFC	LCCCFC	LYCCFC
LFFYY	LCCCCCFC	LCCCBBFC
LCCFFYY	LCFFYY	LCBBYY
LCCCFFYY	LFFLYY	LLFLYY
LFFLLYY	LFFLLLYY	LYYLLLYY
LCFFLYY	LCFFLLYY	CCFFLLYY
LFFYBFYY	LCCFFLYY	LCCFFLFF
LCFFYBC	LFFYBC	LFFYBL
LCCFFYBC	LFFLYBC	YFFLYBC
LFFLLYBC	LCFFLYBC	LCBBLYBC
BLLYY	BLYY	BFYY
BLLLLYY	BLLLYY	YLLLYY
BYBFYY	BLLLLLYY	BLLLLLYB
BLLYBFYY	BLYBFYY	BLYBBYY
BYBFLLYY	BYBFLYY	BCBFLYY
BYBC	BLYBFLYY	BCYBFLYY
BLLYBC	BLYBC	BLYBF
BLLLLYBC	BLLLYBC	BFLLYBC
BLYBFYBC	BYBFYBC	FYBFYBC
BYBFLYBC		

APPENDIX E (CONTINUED)
LETTER STRINGS USED IN EXPERIMENT 3: SET 3

HJJZJ	HJZJ	HHZJ
HJJJJZJ	HJJJZJ	HMJJZJ
HZZMM	HJJJJJZJ	HJJJNNZJ
HJJZZMM	HJZZMM	HJNNMM
HJJJZZMM	HZZHMM	HHZHMM
HZZHHMM	HZZHHHMM	HMMHHHMM
HJZZHMM	HJZZHHMM	JJZZHHMM
HZZMNZMM	HJJZZHMM	HJJZZHZZ
HJZZMNJ	HZZMNJ	HZZMNH
HJJZZMNJ	HZZHMNJ	MZZHMNJ
HZZHHMNJ	HJZZHMNJ	HJNNHMNJ
NHHMM	NHMM	NZMM
NHHHHMM	NHHHMM	MHHHMM
NMNZMM	NHHHHHMM	NHHHHHMN
NHHMNZMM	NHMNZMM	NHMNNMM
NMNZHHMM	NMNZHMM	NJNZHMM
NMNJ	NHMNZHMM	NJMNZHMM
NHHMNJ	NHMNJ	NHMNZ
NHHHHMNJ	NHHHMNJ	NZHHMNJ
NHMNZMNJ	NMNZMNJ	ZMNZMNJ
NMNZHMNJ		

LETTER STRINGS USED IN EXPERIMENT 3: SET 4

GWWRW	GWRW	GGRW
GWWWWRW	GWWWRW	GQWWRW
GRRQQ	GWWWWWRW	GWWWDDRW
GWWRRQQ	GWRRQQ	GWDDQQ
GWWWRRQQ	GRRGQQ	GGRGQQ
GRRGGQQ	GRRGGGQQ	GQQGGGQQ
GWRRGQQ	GWRRGGQQ	WWRRGGQQ
GRRQDRQQ	GWWRRGQQ	GWWRRGRR
GWRRQDW	GRRQDW	GRRQDG
GWWRRQDW	GRRGQDW	QRRGQDW
GRRGGQDW	GWRRGQDW	GWDDGQDW
DGGQQ	DGQQ	DRQQ
DGGGGQQ	DGGGQQ	QGGGQQ
DQDRQQ	DGGGGGQQ	DGGGGGQD
DGGQDRQQ	DGQDRQQ	DGQDDQQ
DQDRGGQQ	DQDRGQQ	DWDRGQQ
DQDW	DGQDRGQQ	DWQDRGQQ
DGGQDW	DGQDW	DGQDR
DGGGGQDW	DGGGQDW	DRGGQDW
DGQDRQDW	DQDRQDW	RQDRQDW
DQDRGQDW		

APPENDIX F
LETTER PAIRS USED IN EXPERIMENT 3: SET 1

PP: 0	PT: 4	PK: 2	PX: 3	PS: 5
TP: 0	TT: 7	TK: 6	TX: 2	TS: 4
KP: 8	KT: 0	KK: 6	KX: 0	KS: 0
XP: 0	XT: 3	XK: 6	XX: 5	XS: 1
SP: 0	ST: 0	SK: 0	SX: 4	SS: 4

REFERENCES

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. Journal of Experimental Psychology: General, 120, 3-19.

Berry, D. C., & Dienes, Z. (1991). The relationship between implicit memory and implicit learning. British Journal of Psychology, 82, 359-373.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). New York: Wiley.

Brewer, W. F. (1974). There is no evidence for operant or classical conditioning in human subjects. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (pp. 1-42). Hillsdale, NJ: Erlbaum.

Broadbent, D. E. (1991). Recall, recognition, and implicit knowledge. In W. Kessen, A. Ortony, & F. I. M. Craik (Eds.), Memories, thoughts, and emotions: Essays in honor of George Mandler (pp. 125-134). Hillsdale, NJ: Erlbaum.

Broadbent, D. E., FitzGerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. British Journal of Psychology, 77, 33-50.

Brody, N. (1989). Unconscious learning of rules: Comment on Reber's analysis of implicit learning. Journal of Experimental Psychology: General, 118, 236-238.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). Journal of Experimental Psychology: General, 120, 316-323.

Dienes, Z. (1992). Connectionist and memory array models of artificial grammar learning. Cognitive Science, 16, 41-79.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. Journal of Experiment Psychology: Learning, Memory, and Cognition, 17, 875-887.

Dukes, W. F., & Bevan, W. (1967). Stimulus variation and repetition in the acquisition of naming responses. Journal of Experimental Psychology, 74, 178-181.

Dulany, D. E. (1991). Conscious representation and thought systems. In R. S. Wyer, Jr., & T. K. Srull (Eds.), Advances in social cognition, Vol. 4 (pp. 97-120). Hillsdale, NJ: Erlbaum.

Dulany, D. E., & Carlson, R. A. (1985). Syntactical judgments and rules: A computer simulation for the separation of guessing of rules from control by rules. Psychological Documents, 15, 19.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? Journal of Experimental Psychology: General, 113, 541-555.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: A reply to Reber, Allen, and Regan. Journal of Experimental Psychology: General, 114, 25-32.

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Estes, W. K. (1986). Array models for category learning. Cognitive Psychology, 18, 500-549.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 234-257.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155-170.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogial transfer. Cognitive Psychology, 15, 1-38.

Harlow, H. F. (1949). The formation of learning sets. Psychological Review, 56, 51-65.

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. Cognition, 28, 249-276.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition of classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-328.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. Cognitive Science, 13, 295-355.

Jacoby, L. J. (1991). A process dissociation framework: Separating automatic and intentional uses of memory. Journal of Memory and Language, 30, 513-541.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. Psychological Science, 3, 172-179.

Knowlton, B. J., & Squire, L. R. (in press). The information acquired during artificial grammar learning. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Kolodner, J. (Ed.). (1988). Case-based reasoning: Proceedings of a workshop on case-based reasoning. San Mateo, CA: Morgan Kaufmann.

Lewicki, P., Hill, T., & Bizot, H. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. Cognitive Psychology, 20, 24-37.

Lewicki, P., Hill, T., & Sasaki, I. (1989). Self-perpetuating development of encoding biases. Journal of Experimental Psychology: General, 118, 323-337.

Lindsay, R. K. (1991). Symbol-processing theories and the SOAR architecture. Psychological Science, 2, 294-302.

Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 539-555.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87, 252-271.

Mathews, R. C. (1990). Abstractness of implicit grammar knowledge: Comments on Perruchet and Pacteau's analysis of synthetic grammar learning. _Journal of Experimental Psychology: General_, _119_, 412-416.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, Jeung, & Druhan, B. (1989). Role of implicit and explicit processing in learning from examples: a synergistic effect. _Journal of Experimental Psychology: Learning, Memory, and Cognition_, _15_, 1083-1100.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. _Memory & Cognition_, _13_, 469-475.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processing in cascade. _Psychological Review_, _86_, 287-330.

Medin, D. L. (1989). Concepts and conceptual structure. _American Psychologist_, _44_, 1469-1481.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. _Psychological Review_, _85_, 207-238.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. _Cognitive Psychology_, _20_, 158-190.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. _Psychological Review_, _95_, 183-237.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. _Journal of Experimental Psychology_, _90_, 227-234.

Millward, R. B. (1980). Models of concept formation. In R. E. Snow, P. A. Federico, & W. E. Montague (Eds.), _Aptitude, learning, & instruction, Vol. 2: Cognitive process analyses of learning and problem solving_ (pp. 245-275). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. _Journal of Experimental Psychology: General_, _115_, 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 700-708.

Neuringer, A., & Voss, C. (1993). Approximating chaotic behavior. Psychological Science, 4, 113-119.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition (pp. 41-82). New York: Wiley.

Perruchet, P., Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. Quarterly Journal of Experimental Psychology, 44A (2), 193-210.

Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situation. Cognitive Psychology, 22, 493-516.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General, 119, 264-275.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. Journal of Experimental Psychology, 83, 304-308.

Reber, A. S. (1967). Implicit learning and artificial grammars. Journal of Verbal Learning and Verbal Behavior, 5, 855-863.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. Journal of Experimental Psychology, 81, 115-119.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. Journal of Experimental Psychology: Human Learning and Memory, 2, 88-94.

Reber, A. S. (1989). Implicit learning and tacit knowledge. Journal of Experimental Psychology: General, 118, 219-235.

Reber, A. S. (1990). On the primacy of the implicit: Comment on Perruchet and Pacteau. Journal of Experimental Psychology: General, 119, 340-342.

Reber, A. S. (1993, May). How to differentiate implicit and explicit modes of acquisition. Paper presented at the 25th Carnegie Symposium: Scientific approaches to the question of consciousness, Pittsburg, PA.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. Cognition, 6, 189-221.

Reber, A. S., Allen, R., & Regan, S. (1985). Syntactic learning and judgments: Still unconscious and still abstract. Journal of Experimental Psychology: General, 114, 17-24.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes of learning a complex rule structure. Journal of Experimental Psychology: Human Learning and Memory, 6, 492-502.

Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. American Psychologist, 45, 1043-1056.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Sanderson, P. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation, and mental models. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 729-747.

Schacter, D. L. (1987). Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 501-518.

Servan-Schreiber, E., & Anderson, J. (1990). Learning artificial grammars with competitive chunking. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 592-608.

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Turner, C. W. (1992, March). An exemplar-based model of an artificial-concept acquisition task. Paper presented at the Southeastern Psychological Association convention, Knoxville, TN.

Turner, C. W., & Fischler, I. S. (1993). Speeded tests of implicit knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1-13.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-357.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 328-344.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. Journal of Experimental Psychology: General, 122, 227-248.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 41, 67-85.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 1047-1060.

BIOGRAPHICAL SKETCH

Carl Turner is a third-generation Floridian who is happy to be leaving before the entire state sinks beneath the weight of poorly built condominiums and monstrous theme parks. He has attempted careers as a parking-lot attendant, rock music disk jockey, and computer programmer, with varying degrees of success. Upon retirement, he plans to build a house on the side of a mountain and practice meditation, perhaps in Austria.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ira Fischler, Chairman
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jeff Farrar
Associate Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard Griggs
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

C. Michael Levy
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Howard Beck
Assistant Professor of Agricultural Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Psychology in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1993

Dean, Graduate School